Smaller Classes or Larger

A Study of the Relation of Class-Size to the Efficiency of Teaching

P. R. STEVENSON
BUREAU OF EDUCATIONAL RESEARCH
OHIO STATE UNIVERSITY

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in the
Faculty of Philosophy, Columbia University

PUBLIC SCHOOL PUBLISHING COMPANY
BLOOMINGTON, ILLINOIS
1923

CONTENTS

		PAGE
	Editor's Introduction	5
	Introduction	7
CHAPTER		
I.	Previous Investigations	9
II.	Existing Conditions and Some Opinions	26
III.	A Study of Class-size in the Elementary School	42
IV.	General Relation between Class-size and the Improvement of Pupils in Elementary-School Subjects	59
V.	Relation between Class-size and Improvement of Bright, Average, and Dull Pupils	70
VI.	The Relation between Class-size and the Improvement of Pupils in an Equal Number of Small-Large and Large-Small Classes	78
VII.	A Study of Class-size in High School	98
VIII.	The Efficiency of Large and Small Classes when Records of Pupils Paired According to Equal Intelligence Scores Are Used	112

243318

Digitized by the Internet Archive
In 2018 with funding from
This project is made possible by a grant from the Institute of Museum and Library Services as administered by the Pennsylvania Department of Education through the Office of Commonwealth Libraries

https://archive.org/details/smallerclassesor00stev

EDITOR'S INTRODUCTION

It is a matter of common knowledge that in different communities there are considerable differences in the sizes of public school classes. One is led to wonder whether there is an equal variation in the results of teaching, and particularly whether in the variations of classes and results there is any parallelism. Rice found no better spellers, no greater ability to solve problems, and no greater mastery of language in the cities where classes were small than in those where classes were large. This seemed paradoxical. Moreover, it was a challenge. Accordingly investigators began, at about the time of Doctor Rice's reports, to give serious attention to the question of size of class.

This went on until about 1915 when the energy devoted to the derivation and standardization of test materials created an interlude of about five years in the investigation of administrative and instructional questions. The interlude, however, had one important result. It placed in the hands of those who desired to resume the study of such questions a powerful instrument.

As interest in administrative and instructional investigation was resumed, the question of size of class again came into its own. The study which is here reported was begun in September, 1920. While it was in progress, three or four other studies on the same subject appeared—enough to suggest a sort of renaissance of interest. One of these—the one by Professor C. O. Davis in the June, 1923, issue of *School Review*—is particularly important, not so much because of refinement of methods as because of the auspices under which the investigation was conducted.

So far, however, the investigators in this field have raised as many problems as they have attempted to solve. It is a worthy field of study, for the number of students per class enters powerfully into the determination of the cost of instruction—more so, for example, that teachers' salaries or their hours of teaching as these are generally administered.

How potent this matter of class-size is and how much it may be at variance with the visible product of instruction may be illustrated by the group of forty-four classes in English which Doctor Stevenson describes in his seventh chapter. These classes were arranged in pairs, one member of the pair being large and the other small. Each large class with its corresponding small class was taught by the same teacher. The average size of the large classes was 38 pupils, that of the small classes, 20 pupils. The average term mark of the large classes was 77, that of the small classes, 78. From this we may raise the question of cost in something like the following form. If it costs $15.00 per pupil to teach English to the point represented by a term mark of 77, is it worth $28.00 per pupil to teach it to the point represented by 78? Or, in other words, is it worth $13.00 per pupil to raise the term mark one percentage point?

Perhaps such a question may seem unfair. Teachers' term marks are unreliable, we know. But the unreliability about which we read so much is generally the single rating of a pupil's paper or performance. Term marks, which are presumably based upon many observations, may be expected to be more reliable. Anyway, they are the ratings that count. They are the bases of passing, promoting, and graduating pupils. They are recorded in books and handed around from institution to institution. We cannot ignore them.

The study which Doctor Stevenson reports in this monograph is suggestive at many points. It puts forward a method which, as exhibited in the part of his report devoted to the elementary schools, has promise of high validity. His data, however, are too few and his large and small classes differ too little to permit him to draw rigorous conclusions—as he readily admits. This he is quick to admit. The only thing to do is to use the method on a larger scale and under more favorable conditions. This he is doing. We shall look forward with interest to his next report.

B. R. BUCKINGHAM, *Editor.*

October 10, 1923.

RELATION OF CLASS SIZE TO EFFICIENCY OF TEACHING

INTRODUCTION

In September, 1920, a cooperative arrangement was entered into between the public school system of Chicago and the Bureau of Educational Research of the University of Illinois. According to this arrangement three problems were to be studied—problems in which the administrative offices of the schools of Chicago were interested and in pursuance of which data from the schools of Chicago were to be used. One of these problems involved the investigation of the effect of size of class upon efficiency of teaching. The preliminary plans for pursuing this project were made by Superintendent Peter A. Mortenson for the schools of Chicago and by Doctor B. R. Buckingham for the Bureau of Educational Research of the University of Illinois. The detailed arrangements were handled by Assistant Superintendent Ambrose B. Wight who is in charge of measurement work for the schools of Chicago and by the writer of this report who represented the Bureau of Educational Research of the University of Illinois.

The problem was a practical one which arose in the administration of the public schools of Chicago and in a very real sense this report is an attempt to afford, for the benefit of these schools, an answer to the problem which they found to be confronting them.

Elementary schools were selected by Mr. Wight as representative of the system and four high schools were also selected with the same representative idea in mind. Supplementary data were obtained from four other high schools of Illinois.

It need scarcely be pointed out that the writer is indebted to the school people of Chicago—not only to its administrative offices but also to the participating principals and teachers—for the opportunity which made this investigation possible. Not only

was all the test material purchased by the Chicago Board of Education but a considerable amount of clerical work was also provided by the same agency. The writer visited the schools of Chicago many times, meeting the principals both in general conferences and at their schools, conferring with teachers, and supervising the administering and scoring of tests. He also directed the clerical work, especially that which was done at the University of Illinois. Superintendent Wight arranged for the giving of the tests, took charge of distributing the test materials and collecting them, handled the many administrative details connected with modifying the sizes of the classes, and directed most of the clerical work which was done at the central office of the Board of Education at Chicago. Without his unfailing support and that of his department and without his tact, continually displayed in handling many problems which of necessity arose in the course of the investigation, the experiment could not have been executed with the degree of success indicated by this report.

The writer wishes to express his personal appreciation of the attitude of Superintendent Mortenson in connection with the experiment. He showed from the beginning not only a willingness to support this and other research projects but also a disposition to abide by the results. Moreover, throughout the entire investigation he appealed, with success, to principals and teachers to make such personal sacrifices as might be necessary in order to secure the most reliable data. Having agreed to cooperate, he interpreted his agreement generously.

CHAPTER I

PREVIOUS INVESTIGATIONS

Factors Influencing Results of Teaching

This monograph is the latest of a series of reports which present data on the relationship between class-size and the efficiency of teaching. It recognizes, however, more definitely than do previous reports the fact that in addition to class-size there are several factors which influence the efficiency of instruction such as the intelligence of pupils, the ability of teachers, the spread of ability in the class, the physical equipment of the classroom, the supervision of instruction, and the course of study.

It is very evident that any investigation of class-size which does not eliminate or allow for all of these other factors, cannot lead to a defensible conclusion. If, for example, knowing only the progress of the pupils and the sizes of the classes in which the progress was made, one were to compare a small class in one city with a large class in another, the average intelligence of the two classes might be vastly different; one teacher might be very much better than the other; one class might be taught as one group and the other in two or three groups; the two classes might have the same average ability but one might be made up mainly of average pupils while the other might have a number of very bright and very dull pupils; and one class might have better physical equipment, courses of study, and supervision than the other class. Any of these differing conditions might account in part for the observed progress of the pupils; and under these circumstances it would be impossible accurately to relate the progress to the sizes of the classes.

Several attempts, some of which we shall describe, have been made to measure the relationship between class-size and the efficiency of instruction in the elementary school, but none has been made with reference to the high school. It seems to be taken for granted that relatively small classes are necessary in

the high school. Without the backing of any experimental evidence, the North Central Association of Colleges and Preparatory Schools recommends 25 pupils to a class as a maximum and states that "no recitation class shall enroll more than 30 pupils." The elementary school, however, has no powerful accrediting agency to determine the size of its classes and consequently many elementary classes of 45 to 55 are found in cities where a high-school class of 40 is considered entirely out of the question.

The studies of class-size which have been made with reference to the elementary school have only occasionally utilized standardized tests for the purpose of measuring the achievement of pupils. None of them has utilized the intelligence test as a supplement to the achievement test and, except in one or two minor parts of investigations, no attempt has been made to utilize the principle of the control experiment in attacking this problem. Of course, the chief reason why these recent methods and newer materials have not been employed is because at the time the investigations of class-size were conducted the methods and materials had not been developed. In this investigation, however, not only have subject matter and intelligence tests been used but a definite attempt has also been made to stabilize or allow for all the important factors thought to influence the results of teaching except the one factor of size of class.

Classification of Previous Investigations

It is rather remarkable that no study of the effect of class-size upon the work of the school has appeared during the past five years.* If we leave out of account the work of Dr. J. M. Rice, which was only incidentally a study of class-size, all the investigations which we shall find worth noticing were made between the year 1909 and 1916. Three of these studies, namely those of

*Since this was written two studies have appeared: Almack, John, "Class-Size and Efficiency," *Journal of the National Education Association*, 12:107-9, March, 1923. Davis, Calvin Oliver, "Size of Classes and the Load Accredited by the North Central Association," *School Review*, 31:412-29, July, 1923.

Cornman,[1] Boyer,[2] and Bachman[3] have to do mainly with promotion rates. A fourth study, that of Harlan[4] has to do with promotion rates in part, but it also takes up a number of other considerations. Three other studies, namely those of Rice,[5] Elliott,[6] and Breed and McCarthy[7] are devoted, with progressive degrees of refinement, to the attainment of pupils in various sizes of classes.

Studies of Promotion Rates

Cornman.—Cornman used promotion percents to show the relative efficiency of instruction in large and small classes at Philadelphia. The classes were arranged in three groups as follows: "under 40," "40 to 49," and "50 and over." The promotion percent was determined for each group. The results are given in Table I for the grammar grades, for the primary grades, and for all grades. From the last line of this table it appears that the highest promotion record was made by classes of medium size and that the record of the group of "50 and over" was but 2.9 percent below that of the "under 40" group.

Cornman concludes from his promotion data that: "(1) careful scrutiny of the facts . . . . fails to reveal any advantage in the small classes over classes of medium size as regards promotion percentages; (2) the classes of medium size make, on

---

[1] Cornman, O. P., "Size of Class and School Progress," *Psychological Clinic*, 3:206-212, December, 1909.

[2] Boyer, P. A., "Class-Size and School Progress," *Psychological Clinic*, 8:82-90, May, 1914.

[3] Bachman, F. P., "Final Report of the Committee on School Inquiry," 1:606-609, 1911-1913.

[4] Harlan, C. L., "Size of Class as a Factor in Schoolroom Efficiency," *Educational Administration and Supervision*, 1:195-214, March, 1915.

[5] Rice, J. M., "Scientific Management in Education." (New York, Hinds, Noble, and Eldredge, 1913) Dr. Rice's studies were originally published as a series of articles which appeared in *The Forum* from October, 1892, to June, 1893, inclusive.

[6] Elliott, Charles H., *Variation in the Achievements of Pupils.* New York: Teachers College, Columbia University, 1915, 114 pp. (Contributions to Education, Teachers College, Columbia University. No. 72.)

[7] Breed, Frederick S. and McCarthy, Grace D., "Size of Class and Efficiency of Teaching," *School and Society*, 4:965-71, December 23, 1916.

TABLE I. PROMOTION RATES CONDENSED FROM CORNMAN

	Under 40			40–49			50 and Over		
	Number of classes	Average Number of pupils	Per-cent pro-moted	Number of classes	Average Number of pupils	Per-cent pro-moted	Number of classes	Average Number of pupils	Per-cent pro-moted
Grammar Grades.....	39	36	85.8	60	45	87.3	13	53	89.3
Primary Grades.....	44	36	80.9	116	45	83.0	48	53	77.9
All Grades...	83	36	83.2	176	45	84.5	61	53	80.3

the whole, the best showing; (3) the large classes do not, on the whole, fall much below the other groups; (4) in the grammar grades, the larger the class the better the promotion record; (5) large classes (50 or more pupils) make a poorer showing in the primary than in grammar grades."

Other data secured by Cornman.—To secure additional data Cornman obtained the monthly report cards (for February, 1909) upon which the pupils' conduct and progress in school work was recorded. Analysis of the data concerning progress in school work showed (1) that in the grammar grades the largest classes made the best showing; (2) that in the total for all grades the medium-sized classes made the best record; and (3) that in the primary grades the smallest classes rated highest. The study of pupils' conduct showed (1) that the percent of pupils rated as satisfactory in conduct was greatest in the larger classes; and (2) that the pupils of medium-sized classes were rated lower than the pupils in both the large and small classes. Cornman suggests the maintenance of classes from 40 to 50 and that certain individuals be transferred to special classes of 15 to 25 for individual instruction.

Boyer.—Boyer made a similar investigation in 1913 which included ten districts in Philadelphia. Dividing the classes into six groups increasing by fives from thirty to fifty and over, he then determined the percent of promotion for each group.

The highest percent of promotion in four of the eight grades was found in classes having less than 30 pupils. In three of the remaining grades the highest promotion percent was in classes having 35 to 40 pupils, while the lowest promotion percent for five of the grades was in the largest-sized group, i.e., 50 and over. On combining the four grammar grades, a regular descent in promotion rates was shown as class-size increased, i.e., from 89.8 for the smallest classes to 82.5 for the largest classes. The primary rates, however, indicated that medium-sized classes had the advantage. Boyer suggests that a study using standardized tests would measure the efficiency of teaching in classes of different sizes far better than the method of promotion percents. He makes no attempt, however, to use the better method.

Bachman.—Bachman used methods at New York similar to those of Cornman and Boyer. He found that classes of 50 pupils and under had a slightly higher promotion rate than classes of over 50. In his conclusion he says: "Thus, although the higher rate of promotion is found, in the majority of grades, in classes of 50 and under, this higher rate is so small that, had promotions in each of the several grades been the same for classes of over 50 as for classes of 50 and under, there would have been, in classes of over 50, a net increase of only 789 promotions out of a total of 73,991 pupils—the equivalent of one additional promotion to each 94 pupils in classes of over 50."

Harlan.—As has already been indicated, Harlan dealt with a number of matters; but he also secured data on promotion rates from 1,346 classes in 32 cities. He reached the following conclusions:[8]

"(1) There is no regular decrease in the promotion rate as the size of the class increases.

(2) The maximum promotion rate does not fall uniformly within the smaller class groups, but is found sometimes in the medium and sometimes in the larger class groups.

(3) The minimum promotion rate does, however, with two exceptions, fall within the larger class groups.

(4) In rates for all grades there is slight but not regular decrease as the size of the class increases. The maximum rate for all grades

[8] Op. cit. p. 6.

combined falls in the smallest class group, and the minimum rate falls in the largest class group.

(5) In all the grades except grades II and VII the maximum promotion rate falls in classes of 30 pupils or less.

(6) The promotion rates in all grades are rather uniformly low for classes of 56 or more pupils."

Elliott.—Finally Elliott, in addition to a number of other ways of approaching the problem of class-size, obtained promotion rates from several cities for classes of sizes ranging by fives from 20 or less to over 50. He found no relationship between class-size and promotion rate.[9]

No clear case is made out in these investigations for a relationship between size of class and promotion rate. The evidence is conflicting. Cornman found that classes of 40 to 50 pupils had the highest rates, while Boyer working in the same city found the highest rates in classes under 40. For all grades combined, Harlan found the maximum rate to be in the smallest classes and the minimum rate in the largest. On the other hand, Elliott found no relationship between promotion rates and class-size. The fact is that it is doubtful if promotion rates afford a measure of the extent to which instruction has succeeded in its purpose. The promotion rate is often determined by the general policy of the school. Moreover, teachers tend to set their standards independently of the ability of the class in question. Thus the ability of a large or a small class may be high or low without having much effect upon the promotion rate.

The Relation of Withdrawals to Class-Size

Although it is exceedingly important to know what, if any, influence size of class has upon the number of pupils eliminated from school, only one important study has been made on the question.

Harlan[10] used the same classes and the same methods of grouping as he did in the case of promotion rates. The term

[9] Op. cit. p. 7.

[10] Op. cit. p. 200 ff.

"withdrawal from the class" was used synonymously with "elimination" and included withdrawals for any cause except promotion to another class. Percents were computed on the total enrollment in the class for the different-sized classes of each grade. The percent of pupils permanently withdrawn from the classes of different sizes during the year is shown in Table II.

TABLE II. SHOWING THE PERCENT OF PUPILS PERMANENTLY WITHDRAWN FROM CLASSES OF DIFFERENT SIZES DURING THE YEAR

(After Harlan)

Size of Class	Grades								
	I	II	III	IV	V	VI	VII	VIII	All
10 or less....	*7.7*	*8.6*	*2.8*	16.4	17.5	15.0	15.7	15.0	12.4
11 to 15.....	14.3	8.8	10.6	*9.3*	11.6	*10.6*	13.6	13.1	11.5
16 to 20.....	14.2	9.3	12.5	11.4	*9.8*	10.9	12.0	9.9	11.3
21 to 25.....	14.5	14.1	11.8	11.3	13.2	13.7	13.8	10.1	12.8
26 to 30.....	17.8	14.8	13.3	14.8	13.1	18.1	*9.4*	12.7	14.3
31 to 35.....	18.6	14.8	12.5	10.5	16.2	13.5	15.6	13.8	14.4
36 to 45.....	18.7	19.8	8.6	12.9	15.5	17.4	16.4	*8.6*	14.7
46 to 55.....	22.4	14.8	18.5	13.5	15.2	18.3	9.5	9.7	15.2
56 and up...	24.0	17.5	14.0	28.9	15.8	33.3	18.7	11.1	20.4
All Classes...	16.9	13.6	11.6	14.3	14.2	16.5	13.8	11.6	14.1

The first line of Table II indicates that in grade I in classes with an enrollment of 10 pupils or less, 7.7 percent of all pupils enrolled were permanently withdrawn during the year; in Grade II, 8.6 percent; in Grade III, 2.8 percent; etc. Figures in italics indicate the lowest percent of withdrawals for each grade.

The following facts are evident from Table II:

(1) The lowest percent of withdrawals falls within classes of 20 pupils or less in all grades except VII and VIII; the highest percent falls within classes of 36 or more pupils except in Grades V and VIII.

(2) As the size of the class increases, there is a somewhat regular increase of withdrawals.

(3) A difference of 4 percent is found when the average percent for all classes of 35 pupils or less is subtracted from the

243318

average for all classes of 36 or more. This indicates that in classes of over 35 pupils, 4 more pupils per hundred are eliminated than in classes of less than 36 pupils.

From these facts Harlan concludes that: "(1) The larger the classes, the higher the percentage of permanent withdrawals from those classes, and (2) large classes, if not the cause of, are, at any rate, accompanied by a high percentage of permanent withdrawals from these classes."

Although in estimating the service of the school the amount of elimination is of unquestioned importance, it is doubtful whether we have in Harlan's investigation more than a tentative expression of the relation of withdrawal to size of class. The reasons why children leave school are numerous, and it has not been customary to regard the size of the particular class to which the child belongs at the time of leaving as one of them. The ages of children, their rate of progress in school, the character of the home (especially its economic condition), and the general attractiveness of the school are factors which unquestionably influence elimination; and they are factors which are not allowed for in Harlan's investigation. Moreover, it is probable that the children, especially those in the upper grades, have tended to be taught in classes of various sizes so that the effect of class-size is not revealed when only the size of the class from which the withdrawal took place is observed and reported. This doubtless accounts for the fact that the percents of withdrawals as shown in Table II differ (except in the very largest classes) by such small amounts.

Classroom Management and Class-Size

Four topics, all of which have been investigated by Harlan, may be considered in this connection: (1) the proportion of pupils giving attention to the recitation; (2) the extent to which the pupils participate by reciting or in some way taking part in the activities of the recitation; (3) the time spent in performing routine activities; and (4) the amount of time wasted by the class. The data concerning these four matters were obtained by

personal observations made in two city school systems during seventy-six visits to thirty classes.

Attention to recitation.—The investigator counted the number of pupils in a class evidently not giving attention as much as half of the recitation period. Doubtful cases were credited as giving attention. The percent of pupils not giving attention was computed for each class. It was found (1) that there was no regular increase in the number of inattentive pupils as the classes became larger, and (2) that class-size is not an important factor in determining the number of pupils giving attention to the recitation.

Participation.—The extent to which children who belonged to classes of various sizes failed to take part in the work in hand was ascertained for the same 30 classes as were observed for attention. Only those were counted as non-participants who neither responded when called upon nor volunteered to recite, or who neither asked questions nor in any way took part in the activities of the recitation. The investigator summarizes a distribution of percents of non-participants as follows: "There is no regular increase in the percentage of pupils not taking active part in the recitation corresponding to the increase in the size of class."

Routine activities.—The number of seconds classes required to complete an activity, from the time the signal was given until the completion of the activity, was obtained for each of the 30 classes mentioned above. The activities measured were (1) time to get to work after each intermission, (2) time to get to work after each recitation, (3) time to get to the board or front of the room, and (4) time to get ready for dismissal.

Harlan summarizes his observations by showing that the greatest amount of time required for most of the activities was in classes of from 16 to 20 pupils; that there was no regular increase in the amount of time used in performing these activities as the size of the class increased; and that the observed facts seem "to indicate that size is not an important factor in efficient classroom management."

Time wasted.—The percent of time wasted was found by dividing the sum of the minutes wasted by all individuals by the length of the study period times the number of pupils in the class. Time was considered wasted when the pupil was not doing the assigned work or some other work considered profitable in the opinion of the observer. The time wasted was estimated as accurately as possible for each individual in the class.

Harlan says in regard to this part of the experiment that "although the data . . . . . are inadequate for conclusive proof, the results would seem to warrant these conclusions: (1) The largest percentage of time wasted is in classes of 10 pupils or less. (2) The smallest percentage of time wasted is in classes enrolling from 26 to 35 pupils. (3) Small classes seem to waste more time than large classes."

We do not feel that these matters of classroom management are of great importance in themselves. The important consideration would not be whether children are attentive, whether they participate, or whether time is wasted, unless successful teaching and learning depend upon such matters. Undoubtedly such dependence may be shown; but it would seem to be a better method to measure directly the extent to which learning takes place rather than to measure something which derives its importance only from its relation to learning.

Moreover, the extent to which children give attention, participate, or waste time is decidedly more dependent upon the kind of teacher than upon the size of the class. For example, it is shown by Harlan that a class of 6 pupils wasted 42 percent of its time, while a class of 30 pupils wasted only 2 percent of its time. What right have we to suppose that the low percent in the latter case or the high percent in the former has anything to do with the size of the class? "What would happen if the teachers were interchanged, or if the 6 pupils in the smaller class were interchanged with 6 from the class of 30?"[11]

[11] Smith, Randolph E., "Size of Class as a Factor in Schoolroom Efficiency—A Reply," *Educational Administration and Supervision*, 1:388, June, 1915.

Only with reservations, therefore, may we quote Harlan's conclusion after he has observed promotion rates and classroom conditions and tested the children in arithmetic. "In light of these conclusions the class of medium size (23 pupils) seems too small for the most economical administration of our schools. Small classes are expensive, since they increase the cost per pupil. This added expense does not seem justified when it is known that the differences in achievement of the small over the large classes are small, as has been demonstrated. If one wishes to secure higher promotion rates, higher scores in arithmetic, better attention and wider participation in class work, more efficient classroom management, and better study habits, these things can undoubtedly be secured through improved methods of instruction and more efficient supervision of the larger classes rather than through a reduction in the size of classes."

The Attainment of Pupils in Relation to Class-Size

The feeling is justified that if there is a "best" size of class, the achievement of pupils in the task of learning school subjects should be most favorable in classes of the sizes pronounced best.

Rice's study.—Long before any of the investigations with reference to class-size were made, Rice had made and published his researches into the abilities of pupils in spelling, arithmetic, and language. He did not set up the problem of class-size as such, but he collected his material in such a way as to enable him to relate the achievement of pupils to the sizes of the classes in which they were taught. Six thousand children from the fourth to the eighth grade inclusive were tested in arithmetic.[12] These children were selected from eighteen schools of seven cities. From this investigation Rice concluded that no allowance is to be made for the size of the class in judging the results of his test. He found that the number of pupils per class was larger in the six schools having the highest scores than in the schools having the lowest scores.

[12] Rice, J. M., *Scientific Management in Education*, New York: Hinds, Noble and Eldredge, 1914, Chapter VII.

A study of attainment in language[13] with 8300 children in nine cities in grades four to eight inclusive was also made by Rice. This investigation showed that there was no relationship between results and class-size. Some of the poorest work was done in the smallest classes and some of the best in the largest classes. These investigations were carefully conducted with the tests then available, but nothing is known concerning the relative intelligence of the pupils or the ability of the teachers. Moreover, no information was obtained as to the sizes of classes in which the pupils had been taught prior to those to which they belonged when they were tested. This degree of refinement was only to be found in much later investigations.

Standing in the Courtis arithmetic tests.—Among the different ways by which Harlan approached the problem of class-size was a determination of the ability of pupils in arithmetic. His data were gathered from 675 classes of all grades widely scattered throughout the country. He gets at the effect of the number of pupils in each class by tabulating the number of maximum scores. The test being Series A of the Courtis arithmetic tests, there are 11 scores for each pupil and therefore 11 class scores. Since 6 grades (III to VIII) were tested, 66 maximum scores are possible. The following results are found with respect to these maximum scores when they are distributed among classes ranging by 5's from "10 or less" to "56 and up":

(1) Maximum scores are made in every sized class group.

(2) The largest number of maximum scores is found in the class groups of from 36 to 45 pupils.

(3) The middle 50 percent of maximum scores falls within a range of from 26 to 45 pupils per class.

Harlan also summarizes the test scores on the basis of plus and minus deviations from grade averages. His conclusion on this basis is that (1) little if any superiority in achievement in arithmetic is shown by small classes over large classes, (2) medium-sized classes of 25 to 45 pupils rank slightly above the average for classes in general, (3) almost without exception very

[13] *Ibid.* Chapter X.

large and very small classes do work inferior to classes in general, and (4) without exception medium-sized classes are superior to large and small classes in achievement in arithmetic as measured by the Courtis tests.

The results of this investigation are inconclusive for the same reasons that have been pointed out before. No account is taken of important factors, such as teaching ability, pupil ability, and environmental influences. Moreover, no data are obtained as to the sizes of classes in which pupils have been taught for a series of years.

Elliott's investigation.—In his report on "Variation in the Achievement of Pupils" Elliott made a number of advances in the study of the effect of size of class. Pupils were tested in a larger number of school subjects and some account was taken of other conditions than the size of the class at the date of the testing.

Tests in composition, spelling, range of vocabulary, handwriting, and arithmetic were given to pupils in the fifth and seventh grades of several school systems. In general the following results were obtained: (1) in the fifth grade the larger classes (larger than 35 pupils) were found to be superior in writing, arithmetic, and composition and inferior in range of vocabulary and spelling; (2) in the seventh grade the larger classes were superior in composition and inferior in range of vocabulary, arithmetic, and spelling.

Classes which had been taught in large or small groups for four years were given tests in composition, arithmetic, vocabulary, writing, and spelling. The results showed that (1) the large classes in the fifth grade were superior in composition and arithmetic, and inferior in range of vocabulary, writing, and spelling; (2) in the seventh grade the larger classes were superior in writing and inferior in the other subjects.

Growth Studies

A still further refinement of procedure was adopted by Elliott when he measured the improvement of pupils in writing

and language over a period of seventeen weeks. In one city 83 classes were measured in handwriting with the following results: (1) among classes which showed no growth, the small classes were nearly three times as numerous as the large classes; (2) the lowest percents of growth were made in the small classes; (3) the number of classes making a gain of fifteen percent or more beyond the first measure was greater in the large classes. In 42 classes measured in language, the largest gain in points was made by the larger classes.

From this investigation Elliott concluded that: "Results taken for the sole purpose of studying the improvement made by children show no correlation between class-size and attainment." These negative results are perhaps to be accounted for by the fact that influences more powerful than class-size were allowed to have effect. These influences may have served to mask any clear evidence as to the one aspect of classroom work which was under consideration.

Class-size and efficiency in teaching spelling.—Breed and McCarthy[14] have reported a study of the relationship of class-size and the efficiency of teaching spelling made under the auspices of the Southeastern Michigan Superintendents' Round Table in 1916. The following procedure was carried out for 58 classes:

(1) Classes were arranged in pairs, each pair consisting of one small and one large class of the same grade in the same town or in towns of approximately the same size.
(2) A preliminary test of 80 words was given to each class for the grade to which the class belonged.
(3) After the test the words were studied for 20 school days.
(4) A final test was then given on the 80 words.
(5) The conditions in the two classes compared were as nearly equal as possible:
    (a) The same amount of time was given for recitation and study.
    (b) The same number of words was given each day.
    (c) The same method of teaching was used.
    (d) The same person gave the preliminary and final tests.

[14] Breed, Frederick S. and McCarthy, Grace D., "Size of Class and Efficiency of Teaching," *School and Society*, 4:965-71, December 23, 1916.

(6) In this study ratings of the teachers were secured and an attempt was made to have teachers of equal ability in each of the pairs of classes.

It will be seen from Table III that the large classes show a greater percent of improvement than the small classes except in grades III and VII. This table is based on adjusted records, i.e., on the reports of those pupils who missed no portion of the preliminary and final tests.

Recognizing that the comparison made in Table III might not be reliable, since the classes did not have the same initial ability,

TABLE III. AVERAGE PERCENT OF IMPROVEMENT IN LARGE AND SMALL CLASSES, BASED ON "ADJUSTED RECORDS"

(After Breed and McCarthy)

Grade	Average Percent of Improvement		Differenee in average pereent of improvement	Average Enrollment		Average Number Whose Reports Were Used		Number of pairs of classes
	Large Classes	Small Classes		Large Classes	Small Classes	Large Classes	Small Classes	
3	36.99	37.87	.88−	48.7	30.3	41.0	27.66	6
4	45.73	44.06	1.67+	44.0	26.6	40.5	25.14	7
5	51.67	48.83	2.84+	46.0	26.0	44.33	27.50	6
6	45.79	43.11	2.68+	40.8	25.0	38.0	24.20	5
7	35.92	37.26	1.34−	45.0	22.8	42.80	21.20	5

the investigators sought to compare the classes which had the same initial spelling ability. This was done by pairing the classes in a given grade which in the preliminary test had made 0-9 percent, 10-19 percent, and so on. The average gains of these classes were then computed. By this method equal-ability groups in spelling were obtained in the large and small classes and the improvement of these alone compared.

Substantially the same results were obtained by this method of comparison as were obtained by the method exhibited in Table III. The large classes surpassed the small classes in the

average percent of improvement except in grades IV and VII. On the basis of the number of times a large class surpassed the small class with which it was paired—i.e., upon the basis of the *existence* of greater improvement rather than the amount of this greater improvement—the large classes exceeded the small classes in all grades except the third and fourth. In the third grade, of six pairs of classes three large classes were superior and three small classes. Of the total of twenty-nine pairs of classes, seventeen large classes exceeded the small classes with which they were paired. This amounted to 59 percent of them.

The authors conclude that in amount of improvement computed in either of three ways, the larger classes surpassed the smaller in three of five grades. Continuing, they say, "there was a positive correlation of .07 between size of class and amount of improvement on the list of 58 classes; the correlation increased to .28, probable error .098, for 40 classes ranging in size from 20 to 45 inclusive.

"When classes were arranged according to increasing size of enrollment in groups differing from each other by five, there was a regular increase in the amount of improvement from an enrollment of 20 to an enrollment of 45; after that, a decrease."

The authors recommend that: "Classes in the elementary school of less than 40, especially below the seventh grade, be increased to an enrollment of 40 to 45."

We may sum up all these investigations by observing first, that in general the conclusions do not support the view that small classes are desirable; second, that from the earliest to the most recent studies there has been a progressive refinement of procedure without, however, altering appreciably the first showing to the effect that unless classes exceed 45 or 50, there is no clear indication of diminished efficiency; and third, that even in the latest reports there is a failure to take account of important variables which undoubtedly affect the results of teaching. In the treatment of the question of class-size which we shall set forth in succeeding chapters special consideration has been given to these variables. In other words, the process of refining the

procedure has been carried still further. It will be interesting to observe whether the earlier negative results obtained by cruder methods are supported. In spite of the fact that many refinements have been made in this study over the previous ones, the author wishes to emphasize the fact that this experiment is only a tentative one, and that the conclusions, unless they are supported by other considerations, should not serve as a basis for changing administrative policy in regard to class-size. Before presenting our data we shall examine the actual conditions existing in the American public school in respect to sizes of classes and consider some of the opinions which educational people have expressed. These matters will be discussed in the next chapter.

CHAPTER II

THE EXISTING CONDITIONS AND SOME OPINIONS

In the establishment of educational standards a common method of procedure is to ascertain existing practice. For example, studies of cost are quite generally based on an inquiry into present conditions. Norms in school subjects are likewise expressions of the current abilities of pupils. While the judgment of teachers and of other school people is often pronounced in criticism of the prevailing sizes of classes, it will no doubt be worth showing what these conditions really are.

Sizes of Classes in Chicago

At the beginning of this investigation and as part of it, data were secured from the office of the Chicago Superintendent of Schools on the sizes of classes in a random selection of the elementary schools and in all the high schools. Table IV shows the distribution of classes as thus obtained. The distribution is not only according to size but also according to grade. It is evident that a rather sharp distinction is made between the elementary and high school on the basis of class-size. There were no classes in the elementary schools in which fewer than 15 pupils were enrolled. Indeed, among more than 3,000 elementary-school classes reported, there were but a few straggling classes whose enrollments were less than 30. These amounted to but 31 classes, or about one percent of the total. Sporadic local conditions doubtless accounted for all of them. At any rate, it is certain that the policy of the school administration was not reflected by so small a proportion of the classes. Only 42 of the elementary-school classes had enrollments as small as 30–35, and almost half of these were in the first grade. It is not until we reach the statistical class "35–39" that our frequencies become sufficiently marked to mean much in the way of a tendency. The modal class for the elementary school is "45–49." Not only are

TABLE IV. SIZE OF CLASSES IN CHICAGO SCHOOLS, OCTOBER, 1920

Grade	Number in Class											Total	Median
	Less than 10	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55 and over		
1			3	1	6	20	40	120	248	87	36	561	46.8
2			1	2	1	6	11	102	247	38	1	409	46.7
3			1	1	1	3	18	96	211	48	3	382	46.7
4				1	2	3	15	90	207	39		257	46.6
5			1			2	4	79	215	48	1	350	47.1
6			3	3	1	2	12	103	183	27	4	338	46.2
7					2	2	15	92	180	32	3	326	46.4
8				1		4	54	79	117	15	48	318	44.8
Total Elementary			9	9	13	42	169	761	1618	334	96	3041	46.35
9	36	108	224	556	613	850	683	254	98	92		3474	32.34
10	24	84	202	342	378	401	283	136	17	69		1936	29.27
11	14	60	90	178	148	189	119	31	10	19		858	27.83
12	15	27	64	67	81	93	77	29	2	3		458	28.35
Total High School	89	279	580	1143	1220	1533	1162	450	87	183		6726	30.15

the frequencies in this class greater than in any other, but they are more than twice as great as in any other. Considerably more than 50 percent of all the elementary-school classes reported were in this class—i.e., they had enrollments of 45 to 49 pupils inclusive.

No clear distinction between the sizes of classes in different grades of the elementary school is observable at Chicago. Indeed the outstanding fact in Table IV is that the central tendencies for all grades are substantially the same. Little support is here given to the prevalent idea that in practice the largest classes are in the lowest grades.

In the high school the situation is considerably different. Whereas no classes of less than 15 were reported in the elementary school, and whereas a negligible number of elementary classes were less than 30, there were 368 out of 6,756 high-school classes with less than 15 pupils and 3,311 with less than 30 pupils. In other words, almost half of the high-school classes consisted of fewer than 30 pupils. The modal statistical class for high schools is "30–34," although the concentration of frequencies in this class is by no means as sharp as is the concentration of frequencies in the class "45–49" for the elementary schools. No classes in the high schools were reported as having enrollments greater than 54. On the other hand, 96 elementary classes, or more than 3 percent of those reported, had enrollments in excess of 54.

Between the eighth and ninth grades—i.e., between the last elementary and the first high-school grade—a sharp drop in central tendency is observable. Whereas the median size of class for the eighth grade was 45 (to the nearest whole number), the corresponding figure for the ninth grade was 32. Further reductions in the size of high-school classes were apparent in the tenth, eleventh, and twelfth grades. The typical class in these three grades had about 28 pupils.

The sizes of recitation classes in the different high-school subjects are exhibited in Table V. As has been shown in other investigations, the largest classes were in physical education. Some

TABLE V. SIZE OF CLASSES BY SUBJECTS IN CHICAGO HIGH SCHOOLS, OCTOBER, 1920

Subject	Number of Pupils in Class										Total	Median
	Less than 10	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50 and over		
German	2	1	1		3	1					8	20.0
Shop	8	66	94	228	101	16	3				516	21.97
Home Economics	7	27	42	57	54	18	11	3			219	22.9
Mechanical Drawing	22	29	63	136	97	67	31	9	1		455	24.2
Office Practice	3	4	2	11	4	6	7				37	24.3
French	6	14	33	51	35	38	15	8	2	1	203	24.7
Botany and Zoology	6	12	24	46	49	51	22	5			215	27.0
Art	7	22	35	44	44	59	28	11	2		252	27.04
Chemistry	2	5	15	21	19	39	3	1			105	27.5
Geography	1		17	17	19	14	19	8	1		96	28.4
Spanish	2	15	18	28	66	59	41	16	2		247	29.6
Bookkeeping	2	15	14	23	28	27	30	17	1		157	29.35
Physics	1	2	12	29	18	40	22	5	1		130	30.37
Geometry	1	9	16	36	62	79	51	13	3		270	30.57
Stenography	1	9	23	55	53	72	65	17	4		299	30.63
Latin	3	15	20	31	46	52	56	21	1		245	30.77
English	3	15	60	143	228	352	278	85	10	2	1,176	31.97
Algebra	1	2	19	36	73	133	115	27	2		408	32.7
General Science	1		8	14	50	108	65	28	2		276	33.0
Typewriting	1	1	12	36	50	70	74	29	11		284	33.0
History	4	13	27	49	37	92	86	63	3	1	375	33.1
Physiology			1	13	17	36	44	12	1	1	125	34.4
Arithmetic			5	18	25	43	57	24	1		173	34.5
Physical Education	5	3	19	21	42	61	40	47	39	178	455	38.6

of these classes ran as high as a hundred or more pupils. The conditions of instruction make it advisable that drill classes in physical education be excluded from our consideration. The range for the other subjects is from a median of 34.5 for arithmetic to medians of 22 and 23 for shop work and home economics.[1] The sizes of classes for the latter subjects are doubtless conditioned by the amount of equipment which rooms of the usual size will contain. In other words, there is operating in the case of shop work and home economics, and probably also in the case of mechanical drawing (median, 24.2), a special limiting factor. Doubtless this is also true of some other cases. Whether or not these factors are necessary limiting conditions is a question. At any rate, it is certain that some of the classes in these subjects are much larger than the usual equipment contemplates.

The striking fact about Table V is not disclosed by the medians but by the detailed distributions. Although shop work and home economics, for example, have typical classes of 22 and 23, the distributions nevertheless show that both these subjects are taught to classes of exceedingly various sizes. Home economics is a little more variable in this respect than shop work. Seven classes have less than 10 students while 3 have over 40. On the other hand, physical education with 178 classes above 50 nevertheless has substantial frequencies down as far as the statistical class "15–19."

Conditions in Other Illinois Cities and Towns

Partly for purposes of comparison with Chicago conditions and partly to secure data which would presumably be typical of smaller communities, a schedule was sent to all cities and towns in Illinois in which there were 6 or more elementary teachers. This schedule called for the number of elementary teachers having classes of the following sizes: less than 20, 20–29, 30–39, 40–49, and 50 and over. The term "class" was defined as the

[1] This neglects the median of 20.0 for German on account of the small number of cases.

number of pupils in a room under the instruction of a teacher. This information was sought for the three years 1918-19, 1919-20, and 1920-21. Thus an attempt was made not only to learn the existing status as to class-size in the elementary schools but also to discover the recent trend in this respect.

Of the cities reporting, 180 gave data for the three-year period. Table VI shows for these cities that in the elementary grades the median class-size in 1918 was 41.4, in 1919, 43.3,

TABLE VI. SIZE OF ELEMENTARY CLASSES IN 180 ILLINOIS CITIES FOR 1918-1920

Size of Class	1918—1919		1919—1920		1920—1921	
	Number of Classes	Percent of Classes	Number of Classes	Percent of Classes	Number of Classes	Percent of Classes
50 and over....	313	3.3	342	3.5	350	3.36
40–49.....	5134	54.5	6865	70.0	7161	68.83
30–39.....	3264	34.6	1845	18.8	2050	19.70
20–29.....	598	6.3	646	6.6	719	6.92
Less than 20...	113	1.3	110	1.1	123	1.18
Total.....	9422	100.00	9808	100.00	10403	100.00
Median.....	41.4		43.3		43.0	

and in 1920, 43.0. Thus it is clear that between the first and second years of the period there was an appreciable increase in the number of large classes. Although the difference in medians, amounting to about 2 pupils per class, does not seem large, it really indicates a considerable change. The situation in the third year of the period closely resembled that of the second year so far as the typical size of class is concerned. When we examine the distributions, we note that the modal size of class—which in all three years remains in the "40–49" group—is much more clearly defined in the second and third years. In fact, the frequencies in the modal group have risen sharply from about 55 percent of the cases to 69 or 70 percent. This increase has been drawn almost entirely from the "30–39" group. For it will be noted that the percent of classes for this group has been

diminished by substantially the same amount that the percent for the classes between 40 and 49 has been increased. This means, of course, that the percents for the other statistical groups have remained practically stationary. There is no marked tendency in this three-year period for classes of "50 and over," or for classes of "less than 20" to become more or less frequent.

Conditions in the Country as a Whole

Probably the best recent study on a national basis is that which has lately been reported by the National Committee for Chamber of Commerce Cooperation with the Public Schools and the American City Bureau.[2] This report gives data for 429 of the cities of the United States whose population according to the 1920 census was 8,000 or more. This is about half the cities in the country of this size, although the total population of the cities reporting amounts to 70 percent of the total for all the cities of this size.

Table VII shows the distribution of classes of various sizes for all cities reporting. The following types of schools are distinguished: kindergarten, elementary, junior high, and senior high. The kindergarten distribution is considerably more variable than any of the others. There are surprisingly large numbers of both very small and very large kindergartens. While the central tendency (median) is not far different from that of the elementary schools, the difference between the quartiles, which is 18 for the kindergartens and only 11 for each of the other types of schools, brings out the wide dispersion of the kindergarten figures. The typical size of class for the elementary school is 38, for the junior high school 28, and for the senior high school 25.

The writer of the report from which Table VII is taken says at this point: "The relative size of elementary and high-school

[2] *Know and Help Your Schools.* Second Report. An interpretation of Inquiry No. II. Relating to school buildings and grounds, enrollment, and size of class in the national survey of urban public schools. American City Bureau, March, 1921.

TABLE VII. SIZE OF CLASSES, ALL CITIES REPORTING
(*From Know and Help Your Schools—Second Report*)

Size of Classes	Number of Rooms Having Enrollment Indicated in Groups to the Left			
	Kindergarten	Elementary	Junior High or Intermediate	Senior High School
19 pupils or less	320	2860	1480	10763
20–24 pupils	423	3598	1472	10391
25–29 pupils	687	7807	2365	10544
30–34 pupils	703	15229	1718	5354
35–39 pupils	760	20662	983	2005
40–44 pupils	680	18495	498	1149
45–49 pupils	404	11102	228	359
50–54 pupils	459	2415	64	225
55–59 pupils	136	547	47	130
60 or more pupils	120	480	223	433
Total	4992	83195	9078	41353
Median	37	38	28	25
Q_1	29	32	23	19–
Q^3	47	43	34	30

classes in these cities indicates one of two things: either a class of 38 pupils in an elementary school is, from the standpoint of effective teaching, no larger than a high-school class of 25; or there is a pronounced tendency throughout the country to furnish better classroom facilities to the small group of high-school students at the expense of the much larger group of elementary pupils. The truth seems to lie in the last statement." With the point of view expressed in this quotation we confess that we have some sympathy. So far as we are aware, there is no evidence tending to show that the larger classes which prevail in elementary schools are justified from the point of view of classroom management or effective teaching. On *a priori* grounds it would seem that high-school classes of a given size could be taught as effectively as elementary-school classes of the same size.

An analysis of the figures presented in Table VII is given in Table VIII. The purpose of this table is to show the extent to which the sizes of school systems influence the number of pupils per class. Accordingly, distributions of sizes of classes are shown for small, middle, and large cities. Small cities are

TABLE VIII. SIZE OF CLASSES. THREE SIZE GROUPS

(From Know and Help Your Schools—Second Report)

Cities	Less than 20	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60 or more	Total	Median	Q_1	Q_3
Kindergarten														
Small	68	92	114	84	67	43	26	30	12	21	557	30	24	39
Middle	88	163	217	161	85	93	69	40	19	55	990	31	24	42
Large	164	168	356	458	608	544	309	389	105	344	3445	40	32	50
Elementary														
Small	535	796	2349	2886	3233	2347	821	296	133	96	13492	35	29	41
Middle	721	1095	2201	3654	4018	3054	1248	343	78	86	16498	36	30	41
Large	1604	1707	3257	8689	13411	13094	9033	1776	336	298	53205	39	34	44
Junior High School														
Small	421	458	948	552	296	119	65	13	8	15	2895	28	23	33
Middle	274	341	554	291	246	139	37	7	5	7	1901	28	23	34
Large	785	673	863	875	441	240	126	44	34	201	4282	29	22	35
Senior High School														
Small	3580	3321	3384	1121	439	131	62	30	14	38	12120	24	19−	28
Middle	3000	2762	2883	1426	562	271	128	59	21	88	11200	25	19−	30
Large	4183	4308	4277	2807	1004	747	169	136	95	307	18033	26	20	31

those having a population of 8,000 but less than 30,000. Cities of middle size are those having a population of 30,000 but less than 100,000. Large cities are those having a population of 100,000 and more. The important fact to be derived from Table VIII is that as the size of the school system increases, the classes become larger. This is true in every type of school; but cities of small and of middle size do not differ very markedly in the central tendencies. The median for kindergartens is one larger for medium cities than for small cities. Likewise, the median for elementary and for senior high schools is one larger. The tendency, however, is for the medians in the case of the large cities to be rather markedly above those for the cities of small and middle size. Kindergartens jump to a median of 40 and elementary schools to a median of 39. Doubtless the standardizing influence of colleges and other accrediting agencies restricts this upward tendency for the senior high school—a restriction which is apparently communicated through it to the junior high school.

It is evident from Tables VII and VIII that there is a very wide range of practice with respect to classes of every sort and in each size of community. The same diversity is apparent in other tables not quoted here but exhibited in detail in the report to which we are referring. A full quarter of the kindergarten classes in large cities had 50 or more pupils in them,[3] and there are literally thousands of classes in the elementary schools in all sizes of cities which have more than 50 pupils. On the other hand, there are quite as many classes that have less than 20 pupils. Doubtless even if our knowledge of the effect of class-size on teaching was much more satisfactory than it is at present, there would still be great diversity in practice. If, for example, the doctrine should be established that classes of 30 to 35 were of optimum size, we should still be confronted by the practical difficulty that as soon as classes reached 36, we should violate our principles either by permitting them to be too large or

[3] Perhaps these very large classes are taught by a kindergarten teacher who has an assistant.

by dividing them and making them too small—unless we combined classes belonging to different grades, which would very likely be a remedy worse than the disease.

Even among cities which, according to the classification adopted in connection with Table VIII, would be regarded as "large," there was wide variation. Table IX shows the median sizes of elementary and high-school classes for all cities of more than 250,000 population for which these data were given in the American City Bureau report to which we have already referred.[4] Each of these cities is large enough to permit a type

TABLE IX. NUMBER OF PUPILS IN MEDIAN ELEMENTARY AND HIGH-SCHOOL CLASSES IN REPRESENTATIVE LARGE CITIES

(From Know and Help Your Schools—Second Report)

City	Elementary	High School	City	Elementary	High School
New York	41		Cincinnati	38	26
Chicago	42[a]	31	New Orleans	35	20
Detroit	42	22	Minneapolis	41	
Cleveland	37		Kansas City, Mo.	39	30
St. Louis	48	40	Seattle	34	25
Pittsburgh	42	28	Indianapolis	41	
Los Angeles	36		Jersey City	40	22
San Francisco	38		Rochester	30	30
Washington	38	23	Denver	33	
Newark	40	28			

[a]The number of pupils in the median elementary class in Chicago was found to be appreciably higher by the author, being 46.

of organization that will make possible any central tendencies which are thought to be reasonable. From this table it will be seen that the elementary schools of St. Louis are on a distinctly different level with respect to class-size from the schools of Rochester, the difference amounting to 18 pupils per class. St. Louis with median elementary-school classes of 48 is nevertheless reputed to have an excellent school system. Even its high-school classes are unusually large, the median being 40. The position of Chicago, as far as the elementary school is concerned, ranks with that of Detroit and Pittsburgh.[5]

[4] Data for Buffalo were left out. According to the figures given, the median elementary class was 32 and the median high-school class was 60 plus. The latter figure seems so unlikely that we have ventured to disregard it.

[5] Although the American City Bureau report shows a median of 42 for the Chicago elementary schools, our own figures show 46.

According to the figures of Table IX, some of the differences between medians for elementary and high schools are very pronounced. Thus at New Orleans the difference is 15, classes in high schools averaging but 20 in size. At Washington the difference is likewise 15; at Jersey City it is 18, and at Detroit 20. Indeed, at Jersey City and Detroit the typical elementary-school class is almost twice as large as the typical high-school class. On the other hand, Rochester seems to have the same policy with respect to both high-school and elementary classes. The medians in each case are 30.

These observations, and others that might be made almost without end from Tables VII, VIII, and IX, emphasize the wide diversity of practice throughout the country on the important question of class-size—important not only because of the very prevalent idea that efficiency of instruction is related to it but also because no item of school organization is more powerful in its effect upon the cost of public education.

Some Opinions in Regard to Class-Size

If we now turn from a consideration of the actual conditions to a consideration of the conditions which ought to exist, we find for our guidance a great many opinions. Although little, if any, experimental evidence exists in published form other than that which we have cited in Chapter I, and although this evidence tends to indicate that relatively large classes may be taught effectively, the opinion of educational experts and practitioners leans heavily toward the small class as desirable. "All school administrators are agreed that 40 is the super-maximum number of pupils that should under any circumstances be seated in any elementary schoolroom."[6] The recommendations of a survey commission are contained in the superintendent's report of Melrose, Massachusetts, for 1920. The following passage occurs. The Committee believes "that 35 pupils are enough for any teacher and that 30 would be a better number."[7]

[6] Report of the Superintendent of Schools for 1917-1919, Lincoln, Nebraska, p. 89.

[7] Annual report of the school department, City of Melrose, 1920, p. 4.

Evidently echoing the ideas as well as the phraseology of the Lincoln report of 1917-19, the Holyoke, Massachusetts, report for 1920 contains the following statement: "There are 32 classes with a membership of over 40 pupils, the super-maximum number which all educators agree should, under any circumstances, be seated in an elementary schoolroom."

"In order that the very best work may be done, classes in the schools ought not to contain more than from 35 to 40 pupils. When classes are of this size, it is possible for the teacher to give the time and attention to pupils requisite to the achievement of the best results."[8]

In the Salt Lake City Survey, Cubberley, after referring to the fact that the average membership in 1914-15 amounted to 35 in the elementary-school system, says, "These figures for the elementary schools are high, 28 to 30 pupils in average membership being nearer what the elementary schools should average."[9]

In a chapter on buildings Bennett gives the dimensions of a room which "will conveniently accommodate 40 pupils." He gives no other dimensions for, he says, "Larger classes should never be permitted and hence no provision should be made for them."[10] Similarly, Strayer and Engelhardt recognize 40 as the maximum number of pupils for an elementary class. After giving the proper dimensions for an elementary classroom, they say: "Larger rooms . . . . tempt administrators to enlarge classes beyond the maximum of 40 pupils."[11]

In the American City Bureau report already referred to, the following statement is made: "A teacher can work most effectively with a small class. When the class is too large it is difficult to give the pupil the personal attention essential to successful teaching. Obviously the subject taught, the age of the

---

[8] Strayer, George D. Problems in city school administration. Yonkers: World Book Company, 1916, p. 94.

[9] Cubberley, Elwood P. School organization and administration. (Educational Survey Series) Yonkers: World Book Company, 1916, p. 49.

[10] Bennett, Henry Eastman. *School efficiency—a manual of modern school management.* Boston: Ginn and Company, 1917, p. 21.

[11] Strayer, George D. and Engelhardt, N. L. *The classroom teacher at work in American schools.* New York: American Book Company, 1920, p. 356.

pupils, and other factors will determine the desirable size of the class. Teachers generally agree, however, that the largest class in kindergarten and elementary grades should not exceed 35 to 40 pupils, and in junior and senior high schools 20 to 25 pupils."[12]

Addressing a district conference of the New York State Congress of Mothers and Parent-Teacher Associations, Angelo Patri is reported to have said that "a teacher cannot teach 50 pupils successfully. A class of 20 is just right. A class of 25 should be the extreme. A class beyond 50 is an atrocious crime."[13] Another New York principal—this time a high-school principal—says: "We are told by teachers of primary classes that from 35 to 40 pupils are all that can be taught if the interests of the pupils are to be of paramount importance. The teachers of the grammar grades with equal force assert that from 30 to 35 should be the limit for pupils doing that grade of work. In high schools the teachers with equal vehemence assert that classes of from 25 to 30 are all that a teacher can handle successfully, and a much smaller number is preferable."[14]

The above quotations reflect the views of experts and of administrators. As to the point of view of teachers, the following is offered. In 1921 a committee of modern language teachers made a study[15] of the size of modern-language classes in New York City. After citing the fact that the administration had agreed upon 38 as the maximum size of class, the report shows that 10 percent of the classes had been allowed to exceed this maximum. The report continues: "Conditions of this sort are discouraging to pupils and the despair of teachers. Pupils of high-school age cannot be taught *en masse;* they must have individual attention. They cannot learn to pronounce a foreign language unless they have frequent opportunity to recite. In a

[12] *Know and Help Your Schools.* Second Report. An interpretation of Inquiry No. II. Relating to school buildings and grounds, enrollment, and size of class in the national survey of urban public schools. New York: American City Bureau, March, 1921, p. 17.

[13] *The New York Evening Mail,* May 4, 1922.

[14] Felter, William L. "The size of classes," *Educational Review,* 41:418.

[15] "Report of a Committee of New York Teachers of Modern Language," *Bulletin of High Points,* 1921, pp. 18-19.

class of 40, with a period seldom more than 40 minutes long, the chance for an individual recitation every day is very slight." The report of the New York Modern Language Teachers also quotes from the British report called "Modern Studies" (1918) as follows: "It is probable, however, that few (teachers) can manage more than 20 to the best advantage. Twenty-five may be an outside figure; 30 is certainly an extreme limit."

Teachers of English in the secondary schools of New Jersey are quoted in the report of the New York City teachers of modern language as follows: "In such conditions (40 in a class) no teacher, however excellent, can make progress." The final recommendation of the modern language teachers is "that the first, second, and fourth term classes in modern languages be organized with a maximum of 25 and that classes above the fourth term should never exceed a maximum of 35."

In order to secure some systematic information concerning the opinions of administrators on the question of size of class, a questionnaire was sent to all superintendents in the cities of the country having a population of 25,000 or more. These superintendents were asked to indicate their judgment as to the ideal size of class for grades 1 to 3, 4 to 6, 7 to 9, and 10 to 12. Two hundred seventy superintendents replied. Table X gives a distribution of their replies. In the judgment of these men the elementary classes should be larger than those in the junior and senior high school. In other words, in this particular their judgment of ideal conditions rather closely parallels the actual facts. The median judgment is that the senior high-school class should consist of about 25 pupils. This corresponds almost exactly with the number found in the 429 cities reported in the bulletin of the American City Bureau. (Table VII.) According to the central tendency of the replies of the superintendents, the ideal for elementary classes would be about 6 pupils less than the actual conditions as found in the 429 cities. (Median actually found, 38. Ideal median, about 32.) The superintendents apparently think that the junior high school should have somewhat

larger classes than it actually does have. (Actual size, 28; ideal size, 31.)

The great range of ideals as exhibited by the superintendents is surprising. Some of them thought that classes of 10 or 15

TABLE X. IDEAL SIZE OF CLASSES AS INDICATED BY 270 SCHOOL SUPERINTENDENTS IN CITIES OF 25,000 OR MORE POPULATION

Score	Grades			
	1, 2, 3	4, 5, 6	7, 8, 9	10, 11, 12
0–19	30	18	9	16
20–24	20	21	48	95
25–29	52	42	58	95
30–34	95	95	104	33
35–39	53	73	35	4
40–44	16	18	15	2
45–49	2	2		1
50–54		1	1	
55–59	1			
Total	269	270	270	246
Median	31.71	32.84	30.96	25.63

were ideal. Not a few, however, located the ideal size above 40. In short, there is not much less variation in the conception of what ideal conditions are than in the conditions as actually found. It is evident that there is enough confusion in thought as well as in practice to justify an intensive study of the grouping of children for purposes of instruction. In the next four chapters we shall consider the question as it relates to the elementary school. Chapter VII will be devoted to high-school conditions.

CHAPTER III

A STUDY OF CLASS-SIZE IN THE ELEMENTARY SCHOOL

In this chapter and the following ones evidence regarding the effect of class-size on teaching will be given for the elementary school and for the high school separately. It is possible that a *general* determination of the extent to which size of class affects teaching is not possible. Here, as in the case of so many other administrative questions, it is likely that there are several answers rather than one answer. There may, for example, be a different answer for the elementary school from that which applies to the high school. Moreover, as far as the elementary school is concerned the data are, almost of necessity, limited to the measurement of tool subjects. It is possible, therefore, that different results would have been obtained had subjects such as geography and history been measured.

Again, with reference to the elementary school, there may be one answer for primary grades and another for intermediate or advanced grades. While it was impossible to obtain sufficient data for all grades to make determinations reliable, it was possible to select grades which should be representative of primary, intermediate, and advanced pupils.

The grades chosen for this purpose were the second, fifth, and seventh. All the data were obtained from the Chicago public schools during the first and second semesters of 1920-21. The particular schools represented were the Washington, Cleveland, Lowell, Farragut, and Corkery. Although a relatively large number of classes began the experiment, completely usable data for the entire year were obtained from a much smaller number. In the second grade 18 classes, in the fifth 24, and in the seventh 8, or a total of 50 classes, carried the experiment through to its conclusion. These 50 classes are the only ones considered in this report.

Method of Procedure

At Chicago semi-annual promotion takes place. Accordingly, at the beginning of the experiment some of these classes were in the lower and others in the upper half of the yearly grades. Those which were in the upper half of the second grade at that time naturally became low third-grade classes the second semester. Similar statements may be made regarding the classes of the high-fifth and high-seventh grades. In a sense, therefore, every yearly grade of the elementary school was represented in the experiment except the first and fourth.

The 50 classes from which data were used were taught by 25 teachers. Each teacher, in other words, taught a pair of classes. One of these classes she taught the first semester and the other the second. On one of these occasions the class was larger than on the other. Sometimes the larger class was organized in the first semester and sometimes in the second. As will be seen later, the difference between the classes in point of size was not as great as would have been desirable from the purely experimental point of view. The experimental point of view, however, was not the only one. The actual conditions under which the experiment was carried on precluded the wide difference between large and small classes which would have been theoretically advantageous. Prior to the beginning of the experiment, the classes in the elementary schools in which the investigation was conducted already had an average of about 45 pupils. Most persons would regard such classes as distinctly "large." When, however, the attempt was made to organize some classes of 25 or 30, the problem of what to do with the rest of the children became difficult to handle. Either more teachers would have been required or other classes would have had to be made larger than the seating capacity of the rooms would accommodate. Under these circumstances the reader will understand how it comes about that the distinction between "large" and "small" classes is not greater.

It is worth mentioning, however, that in actual practice the choice between a policy of very small classes and a policy of

very large classes is seldom possible. For example, no superintendent is likely to be in a situation where the alternative of classes of 20 or classes of 50 is permitted. What really happens is either that classes gradually increase in size as the school population outruns funds and buildings, or that classes gradually decrease because of a consistent policy on the part of the school authorities. In the first case, to arrest the increase, a building policy may be inaugurated, but it will scarcely contemplate the reduction of the existing average size of classes by more than a small number of pupils. A reduction of ten pupils per class would be regarded as large. In the second case, a policy of retrenchment may tolerate larger classes, but it would hardly be based upon an increase of more than ten pupils. It is, therefore, clear that the advantages of large or small classes are in practice confined—so far as large policies are concerned—to advantages to be derived from classes which on the average do not differ greatly in size. Moreover, if there is a real difference between the progress which pupils may be expected to make in large and in small classes, this difference should be appreciable even though the difference between the size of the classes is relatively small. This becomes conspicuously true if, as will be shown later, most of the other influences affecting efficient teaching are either stabilized or allowed for. In other words, under the conditions of the experiment here reported, actual differences ought to be found between results in small and large classes even though the differences may be expected to be small.

Each pair of classes was taught by the same teacher. The two classes consisted of identical pupils except as certain children were added to or subtracted from the first-semester group to vary the size of the class. The data presented in this chapter have reference only to the pupils who were members of *both* first-semester and second-semester classes; or, to state the matter another way, results were used only for children who were taught by the same teacher in both a large and a small class.

Tests Used

Intelligence tests were given to all pupils near the beginning of the school year. "The Dearborn Test of Intelligence," Series I, and the "Indiana Primer Scale" were given to the second-grade pupils. "The Illinois General Intelligence Scale" was given to the fifth and seventh-grade pupils. Pupils added to each class at the end of the first semester were also given the intelligence tests. Care was taken to have the average and variability of intelligence of the classes the same for both semesters. Attainment tests were given to the pupils at the beginning of the first semester (Trial I), at the end of the first semester (Trial II), and at the end of the school year (Trial III). The elapsed time between the first test and the second test was fourteen weeks. The time between the second test and the third test was also fourteen weeks.

The subject matter test used in grade II was the Indiana Scale of Attainment. This test measures spelling, word recognition, arithmetic, and understanding of sentences. Form I was used for Trials I and III, and Form II for Trial II. In grades V and VII the following tests were used: "The Monroe General Survey Scale in Arithmetic," Forms I and II, "The Monroe Standardized Silent Reading Test," Forms I and II, and "The Charters Diagnostic Language Tests," Forms I and II. Form I of each of the above tests was given for Trials I and III, and Form II for Trial II.

Possible Criticisms of Method Used

Here the question of the equivalence of Forms I and II perhaps merits attention. It has been found, for example, that Form II of the Monroe reading test is somewhat easier than Form I. On the other hand, there are data to show that Form II of Monroe's arithmetic test is somewhat more difficult than Form I. Data on the equivalence of the two forms of the Charters language scale are not available, but there is probably an appreciable difference between them. These differences,

however, constitute less of a disturbing element than would at first be supposed. The reading test may be used as an illustration. In the first semester, pupils who take Form I at the beginning of the experimental period and Form II at the end of the semester will register a larger improvement than they are entitled to. On the other hand, in the second semester, these pupils since their initial test for that semester is Form II and since they have obtained relatively high scores on it, will register less improvement than they are entitled to. Now, if large classes prevailed in the first semester and small classes in the second, the result would be to create a bias in favor of the large class.

But the arrangements which were made in the participating schools were such that some of the classes were small the first semester while others were large. Thus we have two different arrangements, (1) A large class in the first semester became a small class in the second semester (consisting largely of the same pupils); this combination will be called "large-small." (2) A small class in the first semester became a large class in the second (again consisting largely of the same pupils); this combination will be called "small-large." In a subsequent chapter we shall consider the amount of improvement in small and large classes when the number of large-small and small-large combinations are equal. By so doing, we shall balance the advantages and disadvantages which accrue to small and large classes in virtue of the fact that the different forms of the achievement tests are not of precisely the same difficulty.

It ought also to be noted that even without the balancing of the two types of pairs, the fact that Form II of the reading test and Form II of the arithmetic test differ in opposite directions from their respective first forms has a tendency to balance the general improvement in small and large classes when all subjects are combined.

A word should perhaps be said as to the practice effect which pupils manifest after they have taken a given type of test. It may be expected that children who have taken Form I of the Monroe reading test will do better on Form II than they would

if they had taken Form II in the beginning. Similarly, and perhaps to an even greater degree, the scores on Form I given at Trial III will be favorably influenced. But these influences are as effective in the case of small-large combinations as in the case of large-small combinations. The reader will bear in mind, therefore, that the differences between scores do not necessarily register the absolute amounts of improvement which took place within the first or second period of fourteen weeks. If the improvement had been measured by three different forms of exact equivalence, and if the practice effect could have been allowed for, we should have the true measure of improvement. The essential idea, however, is the comparative one. If the bias one way or the other is equally observable in both large and small classes according to the combinations we have made, then the relative improvement in large and small classes is determined.

A number of other possible criticisms—criticisms involved in the fact that the large and small classes were taught at different times of the year—may likewise be met by reference to the fact that we are able to arrange our data so that there is an equal number of large-small and small-large combinations. For example, it is a well-known fact that when children return to school in the fall, they have forgotten much that they learned during the previous year. They are out of the school attitude. A test given soon after the opening of school will yield relatively low scores. The immediate improvement, however, is rapid, much of it being due to relearning. This accounts for the fact that in some of our material (especially in the second grade) there is a tendency for large increases to take place in the first semester. Nevertheless, the effect of this bias is neutralized by the arrangement of classes which we have indicated. Again, it may be that the curriculum of the Chicago schools causes material to be taught in one semester or the other which relates more closely to the test material. Such examination as we have been able to give to the offerings of the participating schools does not lead us to believe this condition to be prevalent. In any event, since some of the classes were in the upper half and

others in the lower half of the yearly grades, and since it is possible for us to balance an equal number of each of these against each other, the bias of the curriculum is allowed for.

The essential basis for determining the efficiency of instruction is the improvement registered in terms of the achievement tests within the two fourteen-week periods. Thus it will be seen that the tests given at the end of the first semester (Trial II) constituted the final test with reference to the first semester and at the same time the initial test with reference to the second semester.

At this point one may properly raise the question as to whether this sort of improvement is a sufficient measure of the efficiency of instruction. In other words, it may be objected that standardized tests are not a complete measure of the efficiency of instruction. This objection is very real to a great many people. It is perhaps unnecessary to enter into a lengthy discussion of the question, because in so doing we should be led far afield. It is generally conceded, however, that whatever the purpose of teaching may be, its purpose can only be served to the extent that it produces certain desirable changes in children. These changes may be in the direction of better control of the tool subjects as they are taught in the elementary school and as they are tested by the instruments of measurement. The changes may also be in the direction of improved methods of work, attitudes, appreciations, and ideals. Changes of this type are not directly measured by the standardized tests; but unless there is an opposition between learning the tool subjects and acquiring better attitudes and forms of thought, then it is entirely possible that the achievement tests measure indirectly more of these intangible results of teaching than they measure directly.

The writer will have no quarrel, however, with those who feel that the efficiency of teaching is something more than the ability to secure changes which are registered in the form of higher test scores in tool subjects. Those who take this attitude may simply set as a limitation upon this study of the efficiency

of teaching something which may be phrased as follows: "So far as efficiency is measured or indicated by the tests used."

All the tests were given and scored by the participating teachers. It was quite impossible to provide expert service in the administering of the tests. After the papers had been scored, they were, however, carefully checked either in the superintendent's office at Chicago or at the Bureau of Educational Research of the University of Illinois.[1] Before each testing period the teachers were given detailed instructions in the administering and scoring of the tests either by Doctor B. R. Buckingham or by the writer. The latter also spent considerable time both at the central office and in the schools helping the teachers in administering and scoring the tests.

It should be pointed out here that whatever unreliability may have arisen from the fact that the teachers gave and scored the tests is likewise in part allowed for by the fact that the same teacher tested and scored all the papers of a given child throughout the entire investigation. Any bias which the teacher may have had in one direction or another would be supposed to be equally in evidence throughout the year. Thus, again, our obtained differences between the different trials of the tests may not be the true differences, but they may afford a true basis of comparison. Moreover, all our comparisons are between pupils in one or the other of a pair of classes taught and tested by the same teacher. In other words, the pupils in Miss A's small class are never compared directly with pupils in Miss B's large class. Always the direct comparisons are between Miss A's large and small classes, and (as a distinct comparison) between Miss B's large and small classes.

As has already been intimated, classes were maintained intact as far as possible throughout the year. Among other things

[1]The cooperation of the Bureau of Educational Research and of Assistant Superintendent A. B. Wight in charge of measurements is again noted at this point. It is impossible to state the various ways in which this cooperation worked to the advantage of the investigation. In other words, this is only one of the many instances.

this meant that promotion standards were waived and that all pupils were advanced at the end of the first semester.[2]

Making Scores Comparable

It has been felt to be necessary to adopt some way of converting test scores so that the results for different subjects could be combined. If a child did better in reading while taught in a small class but not so well in arithmetic, the detailed facts are important; but it is also important to know whether on the whole, when reading and arithmetic results are both taken into consideration, the child has benefited more by instruction in the large or the small class.

Again, the whole question of the interpretation of improvements from different initial scores has presented itself. In comprehension of reading, is an improvement from seven to ten equal to an improvement from ten to thirteen or an improvement from fifteen to eighteen? In this test eighteen is the maximum score and it is certain that to improve from fifteen to eighteen is very much harder than to improve three units somewhere lower down the scale. In terms of scale units the improvements are equal; but in the frequency with which such improvements are made, they are not equal.

The task, then, was to devise a plan which would render comparable the scores on the different tests and which would equalize the units from two portions of the growth curve. This question of the growth or improvement from two different positions was particularly important since the same child was measured in the first semester from one position and in the second semester from a different (and usually a higher) position.

---

[2] This need not have been absolutely true in the case of a large-small pair. In such a case some of the pupils in the large class of the first semester would not be members of the corresponding small class of the second semester. If the maintenance of the same average and variability of intelligence did not require the promotion of a given pupil, he might be left back. In any event, he would be transferred to some other class than the small class of this particular pair.

The T-Score Method

The plan adopted for rendering the scores comparable as between tests and with reference to different initial positions involved utilizing the T-score developed by Doctor William A. McCall.[3] This method calls for the making of a distribution table of the scores in a given test, the entering of the corresponding percentile scores, and the transmuting of these percentile scores into the corresponding S. D. values. A T-score unit is one-tenth S. D. The percentile scores are really the reverse of the usual percentile scores. They are the percents of the total number of cases which exceed a given test score. In computing this percent, half the number of cases at the score are counted as exceeding it. In transmuting the percentile scores—or as Doctor McCall calls them, "the percent exceeding plus half those reaching"—into S. D. values, a table is used which represents zero ability as located at a distance minus five S. D. from the mean. On the theory of normal distribution, this is an ability which fewer than one person in a million of a given group, a sampling of which has been measured, will possess.

Table XI shows how the method worked out for the second grade in spelling. Here we find, for example, that 8 children out of 315 spelled 3 words correctly. The number of children who spelled more than 3 words correctly was 303. Adding to 303 half the number of children who spelled 3 words correctly (namely, half of 8) gives 307 as the "number exceeding plus half those reaching." This is 97.1 percent of 315, which gives the entry in the next column of Table XI. To obtain the corresponding T-score any table which expresses the S. D. values corresponding to given percents of the area of the normal surface of frequency may be used, provided these S. D. values are scaled not from the mean, but from a selected zero point at the extreme left of the figure. The table furnished by Doctor McCall[4] is used, the zero being taken, as before stated, at minus

[3] McCall, William A. *How to measure in education.* New York: The Macmillan Company, 1922, pp. 272-306.

[4] *Op. cit.* 274.

five S. D. According to this table a percent of 97.1 corresponds to an S. D. value of 3.1, or a T-score of 31. According to Doctor McCall's plan, T-scores are expressed in units of S. D. when these units have been multiplied by 10.

TABLE XI. DERIVATION OF T-SCORES. SPELLING—GRADE II
(315 Pupils)

Score	Number of pupils	Number exceeding plus half those reaching	Percent exceeding plus half those reaching	T-Score
0	1	314.5	99.5	24.0
1	1	313.5	99.2	26.0
2	2	312	98.7	27.5
3	8	307	97.1	31.0
4	7	299.5	94.8	34.0
5	6	293	92.7	35.5
6	7	286.5	90.7	36.5
7	18	274	86.7	39.0
8	13	258.5	81.8	41.0
9	14	245	77.5	42.5
10	14	231	73.1	44.0
11	16	216	68.4	45.5
12	14	201	63.7	46.5
13	21	183.5	58.0	48.0
14	19	163.5	51.8	49.5
15	10	149	47.2	50.5
16	22	133	42.1	52.0
17	22	111	35.1	54.0
18	20	90	28.5	56.0
19	28	66	20.9	58.0
20	16	44	14.0	61.0
21	11	30.5	9.7	63.0
22	10	20	6.3	65.5
23	12	9	2.9	69.0
24	3	1.5	.5	76.0

Tables similar to Table XI were made for each subject and for each grade. Since there were four subjects in each of the three grades (counting comprehension and rate of reading as separate subjects), there were twelve such tables. It was not felt to be necessary to construct tables of this sort for each of the three trials. Because in almost all cases the second trial of a given test represented for each pupil an intermediate score between the scores made on the first and third trials, it seemed

quite sufficient to construct a T-score table merely from the intermediate trial. In general, therefore, pupils advanced from an initial score below fifty and reached a final score above fifty. The whole scheme can be illustrated by placing pupils on the base line of a probability curve and representing their improvement in terms of T-scores along the base line.

For example, consider the case of a child who secured spelling scores of 8, 12, and 19 on the first, second, and third trials.[5] The median of all second-grade children as obtained from Trial II was 14.5. In virtue of obtaining a score of 8, this pupil exhibited an ability which was surpassed by 81.8 percent of the second-grade pupils according to the data obtained on Trial II. The corresponding T-score as shown in Table XI is 41.0. In Figure 1 the base line, XX′, represents the spread of ability in spelling as exhibited by the data of Trial II. Observe that, as already stated, the median is at 14.5 in terms of words spelled or at 50 in terms of $\frac{\text{S. D.}}{10}$ which is the T-score unit. The placement of this pupil in accordance with his performance on Trial I is at point A (41.0). According to Table XI, his score of 12 on the second trial is exceeded by 63.7 percent of the second-grade children. This places him at a point 46.5 T-score units from zero or at point B. Finally his score of 19 on Trial III places him at point C, a point to the right of which only 20.9 percent of the area of the surface of Figure 1 is found (Table XI). This point is at a linear distance of 58 units from the selected zero point.

Since in Figure 1 the base line, XX′, represents amount of ability scaled from a reasonable zero point, the amount of improvement of the pupil in question for the first semester is the amount by which he moved to the right of point A. In other words, it is the distance AB. Similarly, the improvement made in the second semester is BC.

[5] This is the record of the pupil whose individual score card is shown as Table XII, p. 56.

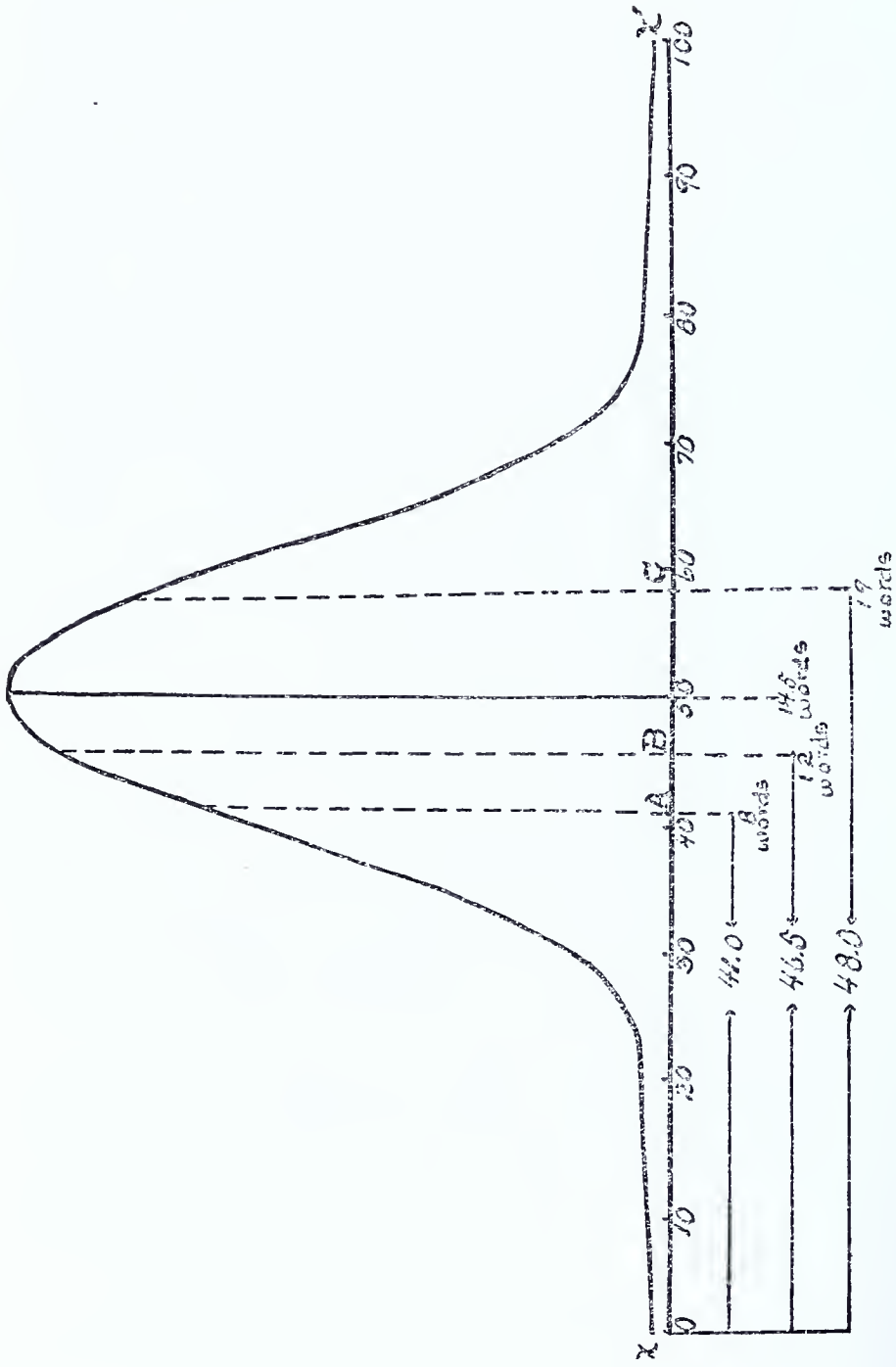

FIGURE I. ILLUSTRATING THE T-SCORE METHOD. DATA REFER TO SECOND-GRADE SPELLING

Justification of T-Score Method

We have said that the device of using T-score units renders our data comparable as between subjects and as regards improvement made from different initial scores. The justification of this statement rests upon the assumption that the distribution of ability in each of the subjects and in each of the grades is governed by the normal curve. With this assumption allowed, we need not even make in this study the further assumption, which is often made in scaling data, that the real variations of abilities in each grade are equal. Whether, for example, the spread (or other function of dispersion) of ability in second-grade spelling equals the spread of ability in second-grade arithmetic or in seventh-grade language is not material in the combining of scores for these subjects and grades. If the variability in seventh-grade language is less, then each unit of variability (each T-score unit, for example) is indeed smaller when referred back to the original test scores; but it is equal to the units derived from distributions of greater variability in the sense that it serves to the same extent to distinguish merit and in the further sense that it constitutes the same proportion of the entire spread of ability of the type in question.[6]

In figures like Figure 1, since the units have the same meaning throughout the entire range, distances expressed by the same number of units will have the same meaning. In other words, in any subject an improvement from 20 to 30 is equal to an improvement from 62 to 72 or from 47.3 to 57.3. Thus the location of the initial score when transmuted into T-score units is immaterial.

Moreover, in terms of T-units a score of 50 in second-grade spelling means the same as a score of 50 in word recognition. In both cases the pupil securing such a score is not only at the median in respect to these abilities but his ability is in both cases at the middle of the range for his grade. Similarly, a T-score

[6] In using the term "entire spread of ability" we take slight liberties with the curve by considering it out off at both ends at a finite distance from the foot of the maximum ordinate—namely, at minus five and plus five S. D.

of 25 in spelling and word recognition means, in both instances, that the pupil receiving such a score is surpassed by the same proportion of his grade and that his status is the same in relation to the range of ability in both spelling and word recognition. This point would be located in Figure 1 and in similar figures one quarter of the distance from X toward X'.[7]

TABLE XII. PUPIL'S INDIVIDUAL SCORE CARD

NAME, *John Smith.* INTELLIGENCE SCORE, *81*
CLASS-SIZE: FIRST, *18* SECOND, *11* GRADE, *5B–5A* CLASS NO., *4*

Test	Trial I	Trial II	Trial III	T-Score Trial I	T-Score Trial II	T-Score Trial III	First Semester Gain	Second Semester Gain	S	L	E
Spelling.....	8	12	19	41	46.5	58	5.5	11.5	x		
Word Recognition.....	6	17	19	32	47	50.5	15	3.5		x	
Arithmetic...	19	31	33	41.5	52	54	10.5	2		x	
Understanding of sentences.....	12	16	19	44.5	52	57	7.5	5		x	

For the reasons stated we have adopted the T-score method of representing the test scores. If our reasoning is valid, we are then permitted to combine any improvements in scores just as we could combine the heights of men or the weights of school children. A fifth-grade child's improvement in rate of reading amounting to 36 T-score units may be related to a second-grade child's improvement in arithmetic of 9 not only by saying that the former is 27 more than the latter but also by saying that it is 4 times as great.

THE PUPIL'S INDIVIDUAL SCORE CARD

The T-score method has been used throughout the discussion of our elementary-school data. The details of our procedure were as follows: Each pupil's score on the different tests was entered on a five by eight card. Table XII shows a pupil's card

[7] Again assuming finite terminations of the curve. Considerations of this sort, however, are practically negligible because of the unusually wide spread of ability provided for by the T-score plan.

containing his raw scores together with the computations derived from them. The first line represents the pupil's record on the three spelling tests. On Trial I this pupil spelled correctly 8 words corresponding to a T-score value (see Table XI) of 41 points; on Trial II he spelled correctly 12 words making a T-score of 46.5 points; and on Trial III he spelled correctly 19 words making a T-score of 58 points. The first term gain of 5.5 is the difference between the T-score of Trial II (46.5) and the T-score of Trial I (41). Similarly, the second term gain of 11.5 is obtained by subtracting the T-score of Trial II (46.5) from the T-score of Trial III (58). The last three columns of the table show by checks or crosses whether the pupil made greater gains in the large or small class or whether equal gains were made.

Since the class to which this pupil belonged had 18 pupils the first semester and 11 pupils the second, his gain in spelling being greater in the second semester was greater in the small class. He made greater gains, however, in the large class[8] than in the small class in all subjects except spelling.

The Class Record Sheet

A second-grade class record sheet for spelling is given in Table XIII. The first line represents the pupil whose detailed record is given in Table XII. The gains made by the other pupils were computed in a similar manner. The reader may obtain a working knowledge of the method of converting raw scores into T-scores by the use of these data in connection with Table XI. In this class the pupils made a total gain of 47 the first semester when taught in a small class and 86 in the second semester when taught in a "large" class. Such class record sheets were made for each grade and for each subject. However, this particular pair of classes was not used in our investigation because the larger class was too small to be treated as a really "large" class. So far as the data in this particular combination are concerned,

[8] The "large" class in this case is not really large since it consists of but 18 pupils. This particular combination was not used in our results. It was selected for illustration because it could be treated briefly. See Table XIII.

TABLE XIII. CLASS RECORD SHEET. SPELLING—GRADE II
(Number of pupils: 1st semester 18, 2nd semester 11)

Pupil	Trial I	Trial II	Trial III	T-Seore Trial I	T-Seore Trial II	T-Score Trial III	First Semester Gain	Seeond Semester Gain	S	L	E
1	8	12	19	41	46.5	58	5.5	11.5	x		
2	14	14	21	49.5	49.5	63	0	13.5	x		
3	13	15	18	48	50.5	56	2.5	5.5	x		
4	12	10	11	46.5	44	45.5	−2.5	1.5	x		
5	9	17	17	42.5	54	54	11.5	0		x	
6	11	7	15	45.5	39	50.5	−6.5	11.5	x		
7	17	19	23	54	58	69	4	11	x		
8	14	13	13	49.5	48	48	−1.5	0	x		
9	14	20	24	49.5	61	76	11.5	15	x		
10	16	23	24	52	69	76	17	7		x	
11	15	18	22	50.5	56	65.5	5.5	9.5	x		
					Total.....		47	86	9	2	

the evidence is that instruction in spelling was much more effective when the class consisted of 11 pupils than when it consisted of 18 pupils.

The Plan of Succeeding Chapters

The method of converting test scores into T-scores and the reason for adopting the method have now been shown. In the following three chapters, which are devoted to the elementary-school part of the investigation, T-scores will be the only scores considered. Moreover, since the T-score unit is equally applicable throughout the entire range for any given subject and grade, the size of a pupil's initial score is of little consequence. His improvement, corresponding to a gain in raw score, is measured by a segment of the base line of the distribution curve. In other words, the length of this segment in terms of T-score units is the amount of improvement or gain. The following chapters will deal largely with these "gains." In a large-small combination, gains in the first semester were made in a large class and gains in the second semester were made in a small class. In a small-large combination, the reverse was true. In either case what we shall be most concerned to know is the extent to which the gains in small classes exceeded those in large classes when they were made by the same pupils and were taught by the same teacher.

CHAPTER IV

GENERAL RELATION BETWEEN CLASS-SIZE AND THE IMPROVEMENT OF PUPILS IN ELEMENTARY-SCHOOL SUBJECTS

The data on which this discussion is based are the records of 305 second-grade pupils, 430 fifth-grade pupils, and 141 seventh-grade pupils. These pupils were taught in 18 second-grade classes, 24 fifth-grade classes, and 8 seventh-grade classes. One semester these pupils were taught in large classes and the other semester in small classes.

Improvement Made by Second-Grade Pupils

Table XIV shows the comparison of gains made in fourteen-week periods by second-grade pupils in large and small classes. All gains are in T-score units. (See Chapter III). It will be seen from columns two and three that none of these classes is what is generally considered small.[1] However, since the average class-size in the Chicago elementary schools is approximately 45, the so-called small classes were small for Chicago. The average difference in size between the large and small classes is twelve pupils. The first line of this table shows the results from one pair of second-grade classes taught by the same teacher. The first semester the average attendance was 50 and the second semester it was 40. Only records from pupils taught and tested in both classes were used. The pupils who made these records were taught first in the class of 50 and afterwards in the class of 40. The sum of the spelling gains made by these pupils when taught in the small class was 298 against 549 made by them in the large class; the sum of gains for word recognition was 258 to 379 in favor of the large class; the sum of gains in arithmetic was 166 to 517 in favor of the large class; and the sum of

[1] In this case as in all cases, the size of a class was determined by using the average attendance as obtained from monthly reports.

TABLE XIV. COMPARISON OF GAINS MADE BY PUPILS IN LARGE AND SMALL CLASSES. GRADE II

Class	Average Number of Pupils		Spelling		Word Recognition		Arithmetic		Understanding of Sentences		All Subjects	
	1st term	2nd term	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
1	50	40	298	549	258	379	166	517	183	462	905	1907
2	49	33	596	325	214	131	147	446	260	318	1217	1220
3	42	51	441	147	439	279	466	349	478	265	1824	1040
4	46	33	349	350	256	315	267	462	273	259	1145	1386
5	36	48	225	223	205	191	221	117	284	146	935	677
6	46	35	403	681	287	151	68	687	75	222	833	1741
7	36	45	568	409	238	126	509	94	303	244	1618	873
8	54	34	239	354	257	303	402	236	242	368	1140	1261
9	48	35	155	438	150	255	62	613	140	612	507	1918
All	..	..	3274	3476	2304	2130	2308	3521	2238	2896	10124	12023
Number of pupils who took all 3 tests			305	305	307	307	298	298	290	290	1200	1200
Average gain per pupil			10.7	11.4	7.5	6.9	7.8	11.8	7.7	10.0	8.4	10.0
Difference in favor of large or small classes												
Amount				0.7	0.6			4.0		2.3		1.6
Percent				6.5	8.7			51.3		29.9		19.0

gains in the understanding of sentences was 183 to 462 in favor of the large class. Thus, in this pair of classes the same pupils made gains amounting to a total of 1,907 points when taught in the large class and 905 points when taught in the small class. This indicates that in this instance approximately twice as great improvement was made in the large class. In the last two columns of Table XIV the sum of gains for all subjects is given. It will be noted that in only three of the nine pairs of second-grade classes did the pupils make greater gains when they were taught in the smaller classes.

The average gain per pupil in the different subjects was in favor of the large classes except in word recognition. The difference in average gains between the large and small classes were: for spelling 0.7 or 6.5 percent[2] in favor of the large classes; for word recognition 0.6 or 8.7 percent in favor of the small classes; for arithmetic 4.0 or 51.3 percent in favor of the large classes; for understanding of sentences 2.3 or 29.9 percent in favor of the larger classes; and for all subjects combined 1.6 or 19.0 percent in favor of the large classes. The differences in arithmetic and in understanding of sentences are large while the differences in spelling and word recogniton are small.

Improvement Made by Fifth-Grade Pupils

The details as to the gains for grade five are shown in Table XV. The difference in the size of these classes ranges from 5 to 14 pupils with an average of 9 pupils. The total gains for all subjects show that the pupils in six of the twelve pairs of classes did better when taught in the smaller groups and that the pupils in the other six pairs of classes did better in the large groups. However, the total gains made by the pupils in the small classes was 12,249 against 9,964 made by the same pupils in the large classes. An inspection of the differences between the average gains of all fifth-grade pupils in the large and small classes will reveal the following: In arithmetic the large

[2] Percentage determinations are in every case computed on the *smaller* gain as a base.

TABLE XV. COMPARISON OF GAINS MADE BY PUPILS IN LARGE AND SMALL CLASSES. GRADE V

Class	Average Number of Pupils		Arithmetic		Reading Comprehension		Reading Rate		Language		All Subjects	
	1st term	2nd term	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
10	45	36	47	353	509	156	491	123	511	233	1558	1045
11	43	37	78	180	106	160	81	172	211	285	476	797
12	42	52	126	262	172	19	120	37	582	706	1000	1024
13	48	35	417	265	422	−23	378	−22	687	35	1904	255
14	44	37	327	547	431	45	355	55	244	825	1357	1472
15	50	36	277	176	70	76	154	76	658	219	1159	547
16	45	35	238	92	192	35	222	−115	272	339	924	351
17	45	33	−17	246	239	191	141	122	212	265	575	824
18	46	35	219	197	9	131	−81	109	610	96	757	533
19	45	50	205	108	184	306	42	511	346	20	777	945
20	48	35	130	248	260	79	284	7	173	553	847	887
21	45	50	392	−32	58	337	−23	447	488	532	915	1284
All	..	..	2439	2822	2652	1512	2164	1522	4994	4108	12249	9964
Number of pupils who took all 3 tests			430	430	430	430	430	430	420	420	1710	1710
Average gain per pupil			5.7	6.6	6.2	3.5	5.0	3.5	11.9	9.8	7.2	5.8
Difference in favor of large or small classes												
Amount				0.9	2.7		1.5		2.1		1.4	
Percent				15.8	77.1		42.9		21.4		24.1	

TABLE XVI. COMPARISON OF GAINS MADE BY PUPILS IN LARGE AND SMALL CLASSES. GRADE VII

Class	Average Number of Pupils		Arithmetic		Reading Comprehension		Reading Rate		Language		All Subjects	
	1st term	2nd term	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
22	40	45	−83	275	40	147	−52	292	334	−10	239	704
23	47	35	360	341	104	183	106	142	269	316	839	982
24	46	35	570	140	473	258	311	−3	683	154	2037	549
25	44	49	1252	140	272	243	−287	252	672	445	1909	1080
All	..	..	2099	896	889	831	78	683	1958	905	5024	3315
Number of pupils who took all 3 tests			141	141	148	148	148	148	141	141	578	578
Average gain per pupil			14.9	6.4	6.0	5.6	.53	4.6	13.9	6.4	8.7	5.7
Difference in favor of large or small classes												
Amount			8.5		0.4			4.1	7.5		3.0	
Percent			132.8		7.1			773.8	117.2		52.6	

classes have an advantage of 0.9 units or 15.8 percent; in reading comprehension the small classes have an advantage of 2.7 units or 77.1 percent; in reading rate the small classes have an advantage of 1.5 units or 42.9 percent; in language the small classes have an advantage of 2.1 units or 21.4 percent; and in all subjects combined the small classes have an advantage of 1.4 units or 24.1 percent. These differences are rather large except in the case of arithmetic.

Improvement Made by Seventh-Grade Pupils

Table XVI presents a similar comparison of gains made by seventh-grade pupils. Again it will be seen that the difference between the size of the classes of each pair is small. Two of the pairs have a difference of five and the other two pairs have a difference of eleven and twelve respectively. The gains made in all subjects combined are larger for the small classes in two of the four pairs. The differences of average gains in favor of the small classes were large in arithmetic and in language, the former being 8.5 or 132.8 percent and the latter 7.5 or 117.2 percent, while the differences in reading comprehension were small being 0.4 or 7.1 percent. The largest difference on a percentage basis (4.1 or 773.8 percent) was found in reading rate in favor of the large classes. This was due to the fact that two of the small classes (Nos. 22 and 24) did not make any gains but on the contrary showed a loss in reading rate. The average gain made in all subjects by all seventh-grade pupils was 3.0 or 52.6 percent in favor of the small classes.

Improvement Made by All Pupils

In order to show the summary of gains made by the pupils in grades II, V, and VII and in these grades combined, Table XVII has been prepared. The gains of second-grade pupils on word recognition have been combined with the language gains of the other grades, and the understanding of sentences with the reading comprehension gains. The understanding of sentences with-

out doubt measures to some extent reading comprehension. Although the word recognition test does not measure the same type of language ability that the fifth- and sixth-grade tests do, it does test to a certain extent the desirable language abilities of second-grade pupils. We obtained no second-grade scores to compare with those in reading rate for the other grades, nor have we records in spelling for grades V and VII to combine with those for

TABLE XVII. SUMMARY OF GAINS MADE BY PUPILS IN LARGE AND SMALL CLASSES. GRADES II, V, AND VII

Subject	Grade II		Grade V		Grade VII		All Grades	
	Small	Large	Small	Large	Small	Large	Small	Large
Arithmetic	2308	3521	2439	2822	2099	896	6846	7239
Reading Comprehension	2238	2896	2652	1512	889	831	5779	5239
Reading Rate			2164	1522	78	683	2242	2205
Language	2304	2130	4994	4108	1958	905	9256	7143
Spelling	3274	3476					3274	3476
All	10124	12023	12249	9964	5024	3315	27397	25302
Difference in favor of large or small classes:								
Amount		1899	2585		1709		2095	
Percent		18.7	25.9		51.6		8.3	

the second grade. In Table XVII, the figures in all columns except the last two have been directly taken from the three preceding tables. Combining the gains made by pupils in grades II, V, and VII in the different subjects we find that: In arithmetic the large classes have a total gain of 7,239 against 6,846 made by the small classes; in reading comprehension the small classes surpass the large classes with gains of 5,779 against 5,239; in reading rate the small classes have a very slight advantage over the large classes, the gains being 2,242 and 2,205 respectively; in language the small classes show gains of 9,256 and the large classes of 7,143; in spelling[3] the large classes reg-

[3] Refers only to second-grade gains.

ister gains of 3,476 against 3,274 registered by the small classes. The summary of the entire table is indicated by the figures 27,397 and 25,302 at the foot of the last two columns. The difference, amounting to 2,095 units, shows the gross superiority of the small classes over the large classes. This superiority amounts to about eight percent. Since 3,488 pupil-records are involved, the average superiority per pupil is 0.6 of a T-score unit.

Percent of Pupils Making Improvement in Small Classes

Since it is rather difficult to apprehend these differences of gains, the number and percent of pupils making greater gains in the small classes than they did in the large classes is given in Table XVIII. This neglects the amount of the excess of gains for small classes and notes merely its presence or absence. If no superiority for either the small or large classes is revealed, the percent of cases in which pupils make greater gains in small than in large classes will approximate 50. To the extent that the percents shown in Table XVIII exceed 50 they indicate superiority for the small classes. To the extent that they fall short of 50, they indicate superiority for the large classes. These statements are strictly true only if equal gains are distributed, half of them being recorded as excesses of gains for the small classes and half of them for the large classes. This has been done in Table XVIII.

Line one of this table, giving the arithmetic results, reads as follows: (1) there were 305 second-grade pupils of whom 41 percent made greater gains when instructed in the small classes than when instructed in the large classes; (2) there were 430 fifth-grade pupils of whom 49 percent made greater gains in the small classes than in the large classes; (3) there were 141 seventh-grade pupils of whom 63 percent made greater gains in the small classes than in the large classes; and (4) there were 876 pupils[4] of grades II, V, and VII combined of whom 48 percent made greater gains in small than in large classes.

[4] Not 876 *different* pupils. The unit is a pupil taught and tested in one subject, or it is a *pupil-record.*

TABLE XVIII. THE PERCENT OF PUPILS MAKING GREATER GAINS IN THE SMALL CLASSES THAN IN THE LARGE CLASSES PLUS ONE-HALF OF THOSE MAKING EQUAL GAINS. GRADES II, V, AND VII

Subject[a]	Grade II		Grade V		Grade VII		All Grades	
	Number of Pupils	Percent making greater gains in small classes[a]	Number of Pupils	Percent making greater gains in small classes[a]	Number of Pupils	Percent making greater gains in small classes[a]	Total number of Pupil Records	Percent of better records in small classes[a]
Arithmetic	305	41	430	49	141	63	876	48
Reading Comprehension	305	45	430	57	148	53	883	52
Reading Rate	...	..	430	47	148	41	578	52
Language	305	52	420	55	141	68	866	57
Spelling	304	48	...	..	...	..	304	48
All	1219[b]	46	1710[b]	55	578[b]	54	3507	52

[a]Plus one-half of those which were equal.
[b]These figures represent the total number of pupil-records in each grade.

As far, therefore, as arithmetic is concerned, pupils did not on this basis do quite as well in the small as in the large classes. This confirms the result for arithmetic as shown in Table XVII. The next four lines give similar data for reading comprehension, reading rate, language, and spelling respectively.

Neglecting the details as to grades, it is clear that in both phases of reading the superiority for the small class is about the same as the inferiority in the case of arithmetic. This also conforms, in a general way, to the results shown in Table XVII where the basis of computation is the size of the differences of gains rather than the existence of such differences. Again, in the case of language there is a definite superiority for the small class, although this superiority, except in the case of the seventh grade, is not impressive. It is well, however, to state that in the seventh grade the difference in size between the large and small class is not great and that the number of records is small. Perhaps we have a real tendency for reading and language work to be better taught when pupils are grouped in small classes. If, however, there is such a superiority it can scarcely be called great unless it can be shown that more pupils benefit by it than Table XVIII suggests.

Finally, in spelling so far as it relates to the second grade, the children did slightly better in the large classes. Again, however, the difference was not great.

Summary

Based on the pupil records in each grade, the last line of Table XVIII discloses that in the second grade the children did somewhat better in the larger classes but that the superiority lay still more definitely with the small class in the fifth and seventh grades. Despite the fact that the seventh-grade records are not all that is to be desired, the indications are that children in higher grades profit rather more by small classes than do children in the lower grades. If further investigation confirms this, the practice of teaching children in larger classes in the lowest

grades would seem to be justified. We are, however, by no means prepared to take this position since in this matter, as in so many others in connection with our investigation, the outstanding condition is not one of difference but rather of indifference. The small and conflicting superiorities one way or the other by small and large classes indicate that both good and poor teaching may be done in both large and small classes. So far, the size of classes, even when important influences affecting teaching are made constant, seems to determine the result only to a very small degree.

CHAPTER V

RELATION BETWEEN CLASS-SIZE AND IMPROVEMENT OF BRIGHT, AVERAGE, AND DULL PUPILS

While we should investigate the general relationship between class-size and the improvement of pupils in school subjects it is also important to know whether the bright, average, or dull pupils may work to the best advantage in large or small classes. This chapter will show the advantage derived by pupils of different intellectual levels when instructed in large or small classes.

Method of Classifying Pupils Into Bright, Average, and Dull Groups

As has been mentioned before, intelligence tests were given to all the pupils who participated in the experiment. The intelligence scores were used in classifying the children of each grade roughly into three groups; the bright, the average, and the dull. The pupils who were below or who just reached the median of the grade below the one to which they belonged were counted as dull, the pupils who reached or surpassed the median of the next grade above the one to which they belonged were considered bright, and the pupils who fell between these medians were regarded as average. For example, the medians on the Indiana Primer Scale for grades I, II, and III are 45.2, 62, and 70 respectively. The second-grade pupils who obtained a score of 45 or less were considered dull, those who received a score of 70 and over were considered bright, and those whose scores were 46 to 69 inclusive were considered average. The grade medians for the Illinois General Intelligence Scale for cities are: grade IV, 40; grade V, 50; grade VI, 65; grade VII, 75; and grade VIII, 88. Fifth-grade pupils were classified as bright if their scores were 65 or more, as dull if their scores were 40 or less, and as average if their scores were anywhere from 41 to 64 inclusive.

In the seventh grade a score of 88 or over was taken to indicate the performance of a bright pupil, a score of 65 or less that of a dull pupil, and a score between 66 and 87 inclusive that of an average pupil. This classification is entirely arbitrary and would not be a valid method of closely classifying individual pupils. However, it does answer the purpose of this investigation since it sharply divides the pupils into three distinct groups. It is evident that a second-grade pupil with an intelligence score of less than the median first-grade pupil is of distinctly lower intelligence than the pupil who reaches or surpasses the median of grade III, or in other words, who is in the upper 50 percentile of the grade next above his own. Likewise, it is clear that among second-grade pupils an intermediate intelligence is shown by those who score between the medians of the first and third grades.

Improvement Made by Different Groups in Grade II

After thus classifying the pupils of the different grades into bright, average, and dull groups, we tabulated the gains they made in large and small classes. Table XIX presents the records of one pair of second-grade classes. The code number given to these classes in Table XIV is four. The reader will note that the total gains for all pupils given in Table XIX checks with those given for class four in Table XIV.

In this pair the first-semester class had an average membership of 46 pupils and the second-semester class an average membership of 33. Of the 32 pupils who were in the classes both semesters four were classified as bright, twenty as average, and eight as dull. The reader will note from this table that the four bright pupils did better in the large classes for each subject tested; that the twenty average pupils made greater gains in spelling and in arithmetic when instructed in the large class, greater gains in understanding of sentences when taught in the small class and equal gains (within one point) in word recognition; and that the eight dull pupils made better records in the small classes for spelling and understanding of sentences and

TABLE XIX. COMPARISON OF GAINS MADE IN ONE PAIR OF SECOND-GRADE CLASSES BY BRIGHT, AVERAGE, AND DULL PUPILS

(Number of pupils: first term, 46; second term, 33)

	Number of Pupils	Spelling		Word Recognition		Arithmetic		Understanding of Sentences		All Subjects	
		Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
Bright	4	37	52	22	53	34	67	24	53	117	225
Average	20	138	219	168	167	169	285	163	151	688	822
Dull	8	124	79	66	95	64	110	86	55	340	339
All	32	349	350	256	315	267	462	273	259	1145	1386

poorer records for word recognition and arithmetic. From the data secured for this class it seems that the pupils benefiting most by the small class were the dull pupils and that the average pupils benefited more than the bright pupils.

Similar computations were made for each of the other second-grade classes as well as for the classes of grades v and vii. The average gain per pupil for the various items from these classes for each grade are presented in Table XX. In the second grade there were 105 bright children, 136 average children, 64 dull children, and a total of 305 children. As in Tables XVII and XVIII where second-grade gains were combined with fifth- and seventh-grade gains, word recognition records were combined with language records and understanding of sentences records with reading comprehension records. The 105 bright second-grade pupils made greater gains in large classes for all subjects except language. However, the difference between the large and small class gains in spelling was small. The gains made by the 136 average pupils were considerably greater in the large classes but the differences in the case of language and spelling were relatively small. The large classes were likewise favored by the 64 dull pupils. Their excess of improvement was less than that of the average pupils and greater than that of the bright pupils.

As in Chapter IV no account is taken of the tendencies for larger gains to be registered in the first semester. However, this does not prevent our making comparisons as to the extent to which large or small classes favored bright, average, or dull pupils. For example, the large second-grade classes might be favorable to all types of children but still we can distinguish degrees of advantage for the different types. The matter of eliminating the bias of the semester advantage will be taken up in the next chapter.

Improvement Made by Different Groups in Grade V

Of the 430 fifth-grade pupils 66 were considered as bright, 152 as average, and 212 as dull. The bright pupils made greater

TABLE XX. COMPARISON OF GAINS MADE BY BRIGHT, AVERAGE, AND DULL PUPILS. GRADES II, V, AND VII

Grade		Number of pupil-records*	Arithmetic		Reading Comprehension		Reading Rate		Language		Spelling		All Subjects	
			Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
VII	Bright.....	105	7.1	11.1	7.5	9 0			7.6	4.9	11.0	11.9	6.6	7.4
	Average.....	136	7.3	11.9	6.9	10.3			7.3	7.7	11.1	11.6	8.2	10.4
	Dull.....	64	9.0	11.6	7.9	8.6			8.1	8.8	9.9	11.3	8.7	9.8
	Total.....	305	7.6	11.5	7.3	9.5			7.6	7.0	10.8	11.5	8.3	9.9
V	Bright.....	66	3.1	4.8	4.5	2.3	3.2	3.0	11.1	11.7			5.4	5.3
	Average.....	152	5.0	7.1	5.3	3.8	3.7	3.9	12.0	10.3			6.5	6.3
	Dull.....	212	7.0	6.7	7.3	3.7	6.5	3.5	12.1	8.9			8.2	5.7
	Total.....	430	5.7	6.6	6.2	3.5	5.0	3.5	11.9	9.8			7.2	5.8
VII	Bright.....	40	7.2	7.7	3.5	7.5	−0.5	7.2	10.8	3.2			5.2	6.5
	Average.....	72	13.0	5.9	4.0	4.5	1.5	3.1	13.7	7.2			8.0	5.3
	Dull.....	36	25.6	5.9	12.8	5.8	−0.2	4.9	17.6	8.3			13.9	6.2
	Total.....	148	14.9	6.4	6.0	5.6	0.5	4.6	13.9	6.4			8.7	5.7
All	Bright.....	211	5.8	8.5	5.8	6.6	1.8	4.6	9.2	6.7	11.0	11.9	6.8	7.5
	Average.....	360	7.5	8.7	5.7	6.4	3.0	3.6	10.5	8.7	11.1	11.6	6.7	7.6
	Dull.....	312	9.5	7.6	8.1	4.9	5.6	3.7	11.8	8.8	9.9	11.3	9.0	6.6
	Total.....	883	7.8	8.3	6.5	5.9	3.9	3.8	10.7	8.3	10.8	11.5	7.8	7.2

*Average number per subject

gains in small classes than in large classes in reading comprehension, reading rate, and also a slightly greater gain in all subjects combined. The average pupils showed greater gains in small classes in two of the subjects, namely, reading comprehension and language. In arithmetic and reading rate there were small differences in favor of the large classes. The small classes showed a slight advantage in the total gains for all subjects combined. The dull pupils made greater gains in the small classes than in the large classes in all subjects. This difference of gains between the small and large classes is very marked in all subjects except arithmetic.

Improvement Made by Different Groups in Grade VII

In the seventh grade there were 148 pupils. Forty of these were regarded as bright, 72 as average, and 36 as dull. The bright pupils made greater gains in the large classes than in the small classes in all subjects tested except language. The gain in favor of the large classes in reading rate was very large while the gain in arithmetic was small. The average pupils made better records in arithmetic and in language when in small classes and better records in reading comprehension and in reading rate when in large classes. The gains in arithmetic and in language in favor of the small classes are much larger than the gains made in reading rate and in reading comprehension in favor of the large classes. Consequently, the total gains for all average seventh-grade pupils are greater for the small classes than for the large classes. The dull seventh-grade pupils made gains in small classes which were much larger than the gains made in large classes except in the case of reading rate where there was a large difference in gains in favor of the large classes. Again, we find that the small classes are more effective for the dull pupils than for the average and bright pupils.

Improvement Made by Different Groups—All Grades

In the second, fifth, and seventh grades combined there were 211 bright pupils, 360 average pupils, and 312 dull pupils, mak-

ing a total of 883. The bright and average pupils made greater gains in the large classes than in the small classes in all subjects except language. The dull pupils made greater gains in the small classes than in the large classes in all subjects except spelling which represents only the gains from the second grade.

TABLE XXI. BRIGHT, AVERAGE, AND DULL PUPILS WHO MADE GREATER GAINS IN THE SMALL CLASSES THAN IN THE LARGE CLASSES

Grade	Number of Pupil-Records			Number of Pupil-Records Which Were Greater in the Small Classes			Percent of Pupil-Records Which Were Greater in the Small Classes		
	Bright	Average	Dull	Bright	Average	Dull	Bright	Average	Dull
II.....	419	544	256	207	242	117	49	45	46
V.....	263	604	843	138	310	483	53	51	57
VII.....	153	282	143	73	155	89	48	55	62
All.....	835	1430	1242	418	707	689	50	50	56

The relative advantage of the large or small classes in the different grades may be readily seen in Table XXI. Here the number[1] and percent of bright, average, and dull pupil records which represented greater gains in the small classes than in the large classes is indicated. The term pupil-record refers to the gain made during one semester in a given subject by one pupil. For example, the gain made during one semester in arithmetic would be considered a pupil-record. Line one of this table gives the second-grade results. There were 419 records of bright children. In 207 records or 49 percent of these cases, the gains made in small classes were greater than those made in large classes. Of the 544 records of average pupils, 242 records or 45 percent of them were better for the small classes than for the large classes. The 256 records of dull pupils were better in the small classes in 117 or 46 percent of the cases. Similar state-

[1] As in Chapter IV one-half the pupils making equal gains in large and small classes were added to this number.

ments may be made as to the fifth and seventh grades. The percents for all grades combined indicate that small classes were in general of greatest advantage to the dull pupils and of less advantage to the average and bright pupils. As a matter of fact since 50 percent indicates indifference, all the advantage of the small class on this basis is brought about by the presence of the dull pupils.

CHAPTER VI

THE RELATION BETWEEN CLASS-SIZE AND THE IMPROVEMENT OF PUPILS IN AN EQUAL NUMBER OF SMALL-LARGE AND LARGE-SMALL CLASSES

In Chapter III several factors were mentioned which might place serious limitations on the study of class-size. The most important of these is the fact that classes which were large or small the first semester may have undue advantage due to the fact that very rapid improvement is generally made in the beginning of a school subject. Especially might this be true in the second grade where the first intensive study of addition and subtraction is begun. Another important item is the fact that the curriculum may cause material to be taught in one semester or the other which relates more closely to the test material. Still other factors to be considered are the possible differences in difficulty of the various forms of the subject-matter test, differences in the practice effect which the pupils manifest after having taken a given type of test, and differences in difficulty of tests at various points on the scale.

The Rotation Method

The purpose of this chapter is to balance these and other factors so that neither the large nor the small classes can have an undue advantage. In order to do this, classes which were large the first semester were paired with an equal number of those which were large the second semester. In this discussion the former type will be called the large-small classes and the latter type the small-large classes. In both the second and fifth grade there were three of each type and in the seventh grade there were two. This method, called the rotation experiment, has been used by Thorndike and others to measure effects of ventilation

and the effect of motion pictures on improvement of school pupils.[1]

To show the improvement made by classes which were large or small the first and second semester Table XXII is given. In grade II the average gain per pupil is larger in all subjects for the classes of the first semester than for those of the second semester. In other words, large gains were made in the first semester irrespective of the size of class. For example: In arithmetic the three small-large classes made an average gain of 11.2 per pupil the first semester and only 5.2 the second semester. In this case the small classes came the first semester. Again, in arithmetic, the three large-small arithmetic classes made an average gain of 15.8 per pupil the first seemster, and only 5.0 the second semester. Here the large classes came the first semester. Similar results were found for the other subjects in this grade.

In grade V, however, no great increase in gains appeared in favor of the first-semester classes. In fact, the reverse is true in respect to reading comprehension, reading rate, and the combined gains for all subjects. In arithmetic and in language the small classes had a decided advantage over the large classes regardless of the semester bias.

In grade VII there is a tendency in favor of the second-semester classes in reading comprehension and in reading rate. In the other subjects and in the average of all subjects combined, the small classes have the advantage irrespective of the semester in which they were taught. From these data it is evident that the semester in which the class occurred determined to some degree the amount of improvement.

---

[1] Thorndike, E. L., Ruger, G. J., and McCall, W. A. "The effect of outside air and recirculated air upon the intellectual achievement and improvement of school pupils," *School and Society*, 7:679-84, May 6, 1916.

Thorndike, E. L., and Ruger, G. J. "The effect of outside air upon the intellectual achievement and improvement of school pupils," *School and Society*, 4:260-64, August 12, 1916.

Lacy, John V. "The relative value of motion pictures as an educational agency," *Teachers College Record*, 20:452-65, November, 1919.

Thorndike, E. L., McCall, W. A., and Chapman, J. C. *Ventilation in relation to mental work*, Teachers College, Columbia University, Contributions to Education, No. 78, 1916. 83 pp.

TABLE XXII. COMPARISON OF AVERAGE GAINS PER PUPIL MADE IN SMALL-LARGE AND LARGE-SMALL CLASSES

Grade	Type of Classes	Number of Classes	Arithmetic		Reading Comprehension		Reading Rate		Language		Spelling		All Subjects	
			Small class	Large class	Small class	Large class	Small class	Large class	Small class	Large class	Small class	Large class	Small class	Large class
II	Small-large	3	11.2	5.2	10.1	6.2			8.1	5.5	11.4	7.2	10.2	6.0
	Large-small	3	5.0	15.8	6.0	13.3			6.7	9.6	8.1	13.5	6.5	13.2
V	Small-large	3	5.4	2.6	3.1	5.0	1.1	7.5	10.7	9.5			5.1	6.1
	Large-small	3	8.3	6.9	7.6	1.3	8.3	0.6	15.7	8.3			9.9	4.3
VII	Small-large	2	16.2	5.8	3.9	4.9	−4.3	6.9	13.8	5.9			7.1	5.9
	Large-small	2	13.5	6.9	8.4	6.4	6.1	2.0	14.0	6.9			10.5	5.6

At this point the rotation method may be explained from the data of Table XXII for second-grade arithmetic. The average gains per pupil for the three small-large and the three large-small groups were combined so as to exhibit all the gains made when classes were small and all the gains when they were large. In the small-large group the pupils made an average gain of 11.2 in the small classes the first semester and 5.2 in the large classes the second semester. In the large-small group, the pupils made an average gain of 15.8 in the large classes the first semester and 5.0 in the small classes the second semester. Combining[2] the average gains per pupil made in the small classes the first and second semesters, we have $\left(\frac{11.2 + 5.0}{2}\right)$ which gives 8.1. Combining in like manner the average gains per pupil made in the large classes of both semesters, we have $\left(\frac{5.2 + 15.8}{2}\right)$ which gives 10.5. These average gains per pupil made by all second-grade pupils of the paired classes in arithmetic will be found on line four in Table XXIII. But since the figures on the "All" lines of Table XXIII are means, they do not check exactly with the unweighted means which may be obtained from Table XXII. The correspondence, however, is very close.

Improvement Made in Large and Small Classes

The average gains per pupil made in paired small-large and large-small classes for the different subjects is given in Table XXIII. The first line of the table indicates the average gains per pupil made in the different subjects by the bright pupils. The differences between the gains made by these pupils in the large classes and in the small classes are less than one except in reading comprehension and in language. The average and dull pupils did better in the large classes than in the small classes, with a slight exception in the case of the dull children in spelling. In

[2] We have obtained the unweighted mean not wishing to carry to our final figures the influence of the fact that the small classes in the first and second semesters may have differed somewhat in size.

TABLE XXIII. THE AVERAGE GAINS PER PUPIL IN PAIRED SMALL-LARGE AND LARGE-SMALL CLASSES

Grade	Number of classes	Number of pupil-records*	Arithmetic		Reading Comprehension		Reading Rate		Language		Spelling		All Subjects	
			Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
II	6	Bright..... 72	8.9	8.1	7.9	9.5			7.5	6.0	9.0	9.2	8.3	8.2
		Average.... 98	7.4	11.5	7.8	9.9			7.4	7.7	10.2	10.9	8.2	10.0
		Dull..... 38	8.3	11.7	8.4	8.9			7.5	9.5	10.8	10.7	8.7	10.1
		All.....208	8.1	10.5	7.9	9.5			7.4	7.4	9.8	10.2	8.3	9.4
V	6	Bright..... 35	3.2	3.2	4.3	0.1	2.6	2.0	12.3	13.4			5.6	4.7
		Average.... 81	6.2	5.5	4.3	4.1	3.3	4.1	11.5	13.0			6.3	6.6
		Dull.....116	8.1	4.1	5.8	4.0	5.2	5.7	13.7	4.8			8.2	4.7
		All.....232	6.6	4.4	5.0	3.4	4.1	4.5	13.0	9.2			7.1	5.3
VII	4	Bright..... 40	6.3	6.8	3.6	7.6	−0.5	7.2	10.0	3.1			5.1	6.4
		Average.... 72	12.9	5.8	4.0	4.5	1.5	3.0	12.9	6.8			7.9	5.1
		Dull..... 36	25.6	5.9	12.8	5.8	−0.2	4.9	17.1	8.1			13.9	6.2
		All.....148	14.9	6.4	6.1	5.7	0.5	4.7	13.8	6.4			8.7	5.8
II, V and VII	16	Bright.....147	7.1	6.8	5.9	6.7	0.9	4.8	9.5	7.1	9.0	9.2	6.8	6.9
		Average....251	8.6	7.1	5.7	6.5	2.4	3.6	10.4	9.2	10.2	10.9	7.5	7.5
		Dull.....190	11.5	5.9	7.6	5.3	3.9	5.5	13.3	6.5	10.8	10.7	9.4	6.1
		All.....588	9.1	7.0	6.3	6.1	2.7	4.6	11.2	7.9	9.8	10.2	7.9	6.9

*Average number per subject.

each subject the combined gain made by all second-grade pupils was greater in large classes than in small classes, except that in language the gains were equal. From the data presented it is evident that the second-grade pupils did better in large classes than in small classes and that small classes were not of any particular advantage to the bright, the average, or the dull pupils.

Although the small classes were of advantage to the fifth grade as a whole, they were especially so for the dull pupils. Average pupils seemed to profit about equally whether they were in large or small clases. In grade VII the small classes showed a greater superiority over the large classes than in either of the other grades. The bright pupils did better in small classes in only one subject, the average pupils in two subjects, and the dull pupils in three subjects. In other words, small classes in the seventh grade proved to be especially helpful to the dull pupils; average pupils likewise benefited by small classes though not to the same degree; while, on the contrary, bright pupils were favored when taught in the larger groups.

Combining the records of the pupils from all the grades and all subjects we find that: (1) the bright and average pupils were not especially benefited by small classes; (2) the dull pupils made the greatest amount of improvement in small classes; and (3) the difference in gains for all pupils made by large and small classes were in favor of the latter by one point (7.9 on 6.9).

Percent of Pupils Making Improvement in Small Classes

The percent of bright, average, and dull pupils who did better in small classes than in large classes is shown in Table XXIV. As in Chapters IV and V one-half the number of those making equal gains in the large and small classes is added to the number who actually made greater gains in the small classes.

In grade II the percents of the pupils making greater gains in the small classes than in the large classes were 50 for bright pupils, 45 for average pupils, 43 for dull pupils, and 47 for all pupils. All but the bright pupils made better records in large

TABLE XXIV. THE NUMBER AND PERCENT OF BRIGHT, AVERAGE, AND DULL PUPILS IN PAIRED SMALL-LARGE AND LARGE-SMALL GROUPS WHO MADE GREATER GAINS IN SMALL THAN IN LARGE CLASSES

Grade	Bright Pupils			Average Pupils			Dull Pupils			All Pupils		
	Number of pupil-records	Pupils making greater gains in the small classes		Number of pupil-records	Pupils making greater gains in the small classes		Number of pupil-records	Pupils making greater gains in the small classes		Number of pupil-records	Pupils making greater gains in the small classes	
		Number	Percent		Number	Percent		Number	Percent		Number	Percent
II	288	144	50	243	110	45	150	65	43	681	319	47
V	140	77	55	321	161	50	463	270	58	924	508	55
VII	153	68	45	286	166	58	143	89	62	582	323	56
All	581	289	49	850	437	51	756	424	56	2187	1150	53

than in small classes. In the case of the bright pupils neither size of class had the advantage.

In the fifth grade, 55 percent of the bright pupils, 50 percent of the average pupils, 58 percent of the dull pupils, and 55 percent of all pupils made greater improvement in the small classes than they did in the large classes. The chances, therefore, seemed to be 11 to 9, or 11 out of 20 that a fifth-grade pupil would do better in the small class.

Small classes were slightly more favorable to seventh-grade pupils than to those of the other two grades. Bright pupils, however, were marked exceptions, the final figure of 56 percent in favor of small classes in the seventh grade being due to the showing of the average and dull pupils, especially the latter. If the difference between the classes of the seventh grade in point of size had been more pronounced, we should feel that the 62 percent of dull children who did better when they were taught in small classes was real evidence as to the pedagogical advantage of instructing the older children of limited mentality in small classes. Even as it is, we feel that the suggestion to this effect is strong.

The records from pupils of all grades indicate that small classes were of practically no benefit to the bright and average pupils since the chances were about even that a given child would do better or not as well when taught in a small class. The small classes, however, really benefited the dull pupils, although not those of the second grade. The summary shows this benefit to be such that 56 out of each 100 dull pupils did better and 44 not so well in the small classes.

Reliability of Results

Table XXV is essentially devoted to the question of reliability. It is found, for example, that in the second grade the difference between the gains made by all large classes and all small classes for all subjects combined is 1.03 T-units. This difference is obtained from mean gains of 9.36 and 8.33 respectively. Two

TABLE XXV. VARIATION AND RELIABILITY OF GAINS AND DIFFERENCES IN THE PAIRED SMALL-LARGE AND LARGE-SMALL CLASSES. GRADES II, V, AND VII

Grade		Arithmetic		Reading Comprehension		Reading Rate		Language		Spelling		All Subjects	
		Classes		Classes		Classes		Classes		Classes		Classes	
		Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
II	Number of pupils	209	209	209	209			208	208	208	208	834	834
	Mean gain per pupil	8.09	10.30	7.95	9.51			7.431	7.43	9.79	10.20	8.33	9.36
	S. D. of gains per pupil	8.26	8.63	6.60	8.85			7.58	7.90	7.05	7.15	7.73	8.23
	Reliability (P. E.) of gains	.38	.40	.31	.41			.36	.37	.33	.34	.18	.19
	Difference of means in favor of small or large classes		2.21		1.56				.001		.41		1.03
	Reliability (P. E.) of differences		.56		.52				.52		.47		0.26

TABLE XXV. *(Continued)*

Grade		Arithmetic		Reading Comprehension		Reading Rate		Language		Spelling		All Subjects	
		Classes		Classes		Classes		Classes		Classes		Classes	
		Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
V	Number of pupils.....	234	234	234	234	234	234	225	225			927	927
	Mean gain per pupil.....	6.61	4.39	4.99	3.39	4.08	4.51	13.00	9.18			7.12	5.33
	S. D. of gains per pupil.....	7.58	7.00	8.20	8.20	9.15	8.79	9.55	10.25			9.40	8.85
	Reliability (P. E.) of gains...	.33	.31	.36	.36	.40	.39	.43	.46			.21	0.20
	Difference of means in favor of small or large classes...	2.22		1.60			.43	3.82				1.79	
	Reliability (P. E.) of differences.....	.45		.51			.56	.63				.29	

TABLE XXV. *(Continued)*

Grade		Arithmetic		Reading Comprehension		Reading Rate		Language		Spelling		All Subjects	
		Classes		Classes		Classes		Classes		Classes		Classes	
		Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
VII	Number of pupils	141	141	146	146	145	145	142	142			574	574
	Mean gain per pupil	14.90	6.36	6.09	5.69	.54	4.71	13.80	6.38			8.75	5.78
	S. D. of gains per pupil	12.70	8.72	10.79	9.87	10.47	8.60	9.66	9.55			12.10	9.30
	Reliability (P. E.) of gains	.72	.38	.60	.55	.59	.48	.55	.54			.34	0.26
	Difference of means in favor of small or large classes	8.54		.40			4.17	7.42				2.97	
	Reliability (P. E.) of differences	.81		.81			.76	.77				.43	

TABLE XXV. *(Continued)*

Grade		Arithmetic		Reading Comprehension		Reading Rate		Language		Spelling		All Subjects	
		Classes		Classes		Classes		Classes		Classes		Classes	
		Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
II V and VII	Number of pupils	584	584	589	589	379	379	575	575	208	208	2335	2335
	Mean gain per pupil	9.14	6.98	6.31	6.13	2.73	4.59	11.20	7.85	9.79	10.20	7.93	6.86
	S. D. of gains per pupil	9.45	8.10	8.94	9.19	9.80	8.32	10.04	9.35	7.05	7.15	9.63	8.85
	Reliability (P. E.) of gains	.26	.23	.25	.26	.34	.29	.28	.26	.33	.34	.13	.12
	Difference of means in favor of small or large classes	2.16		.18			1.86	3.35			.41	1.07	
	Reliability (P. E.) of differences	.35		.36			.45	.38			.47	.18	

problems arise: (1) Is this difference large enough to be significant, and (2) Is it a reliable difference?

The first problem may be approached by comparing the difference with the means from which it is derived. For example, we have in this instance a difference which shows a 12 percent superiority for the large classes. In other words, 9.36 is 12 percent more than 8.33. It is easy to manipulate such measures and to estimate, for example, in how many years a child taught in a large class might, on the basis of these figures, be expected to gain a year over a child of equal ability when taught in a small class. Another way of looking at the same idea might be to raise a question as to how many years more of work per semester would be done by 100 children if taught in large classes when each child so taught learns on the average 12 percent more than he would if taught in small classes. A third speculation might be indulged in by converting the scores into money, and endeavoring to find out how much less it would cost, merely on the basis of progress, to teach children in large classes if on the average they may be expected to do 12 percent more work than they would do if they were taught in small classes. This particular idea of cost would, of course, be independent of the more usual idea derived from a consideration of the larger number of children who receive instruction for the same teaching cost when a greater number are instructed by one teacher.

But such speculations, while attractive, are of value only to the degree that the scores on which they are based are reliable. This raises the second of the two problems stated above. Returning to the second-grade classes where the large classes surpass the small classes by 1.03 T-units, the question is this: Suppose we were to conduct other experiments of the same sort involving equally large groups of equally representative children; what is the likelihood that we should obtain a difference of gains in favor of large classes, and would this difference probably be of about the order of magnitude as 1.03. In other words, the question is that of the reliability of the obtained difference of gains. The present theory as to reliability is carried out on the basis

of ascertaining what the variability of a very great number of similar determinations would be. This variability is usually expressed in terms of PE, which is really a measure of unreliability in the sense that the greater it is, the more unreliable is the value to which it is attached. What, then, is the PE of the obtained difference 1.03?

"The unreliability of a difference between A and B equals the square root of the sum of the square of the unreliability of A and the square of the unreliability of B." This is Thorndike's statement.[1] This unreliability may be expressed in terms of the Standard Deviation, although as has been said it is more usually expressed in terms of PE or of 0.6745 SD. The formula in these terms may be expressed in the following form: $PE_{diff} = \sqrt{(PE_A)^2 + (PE_B)^2}$, where PE_A is the probable error (unreliability) of one of the means and PE_B is the probable error of the other.

Accordingly, the probable error of the two means must first be obtained. The formula for this is $PE_M = \frac{0.6745\ SD_{dis}}{\sqrt{N}}$ In this formula PE_M is the general expression for the probable error of any mean. If we are concerned with the mean of characteristic A, it becomes PE_A; if we are concerned with the mean of characteristic B, it becomes PE_B. In this formula SD_{dis} is merely the Standard Deviation of the distribution of the measures from which the mean in question (either A or B) was derived.

The first step therefore in deriving an expression for the reliabilty of our obtained measure of difference is to compute the Standard Deviation of the individual gains of pupils. This has been done for each grade and subject for both the large and small classes considered in this chapter. The results are entered in Table XXV on the line called "SD of gains per pupil." Using these SD's, and the formula for PE_M, the reliability of each mean gain has likewise been computed and entered in Table XXV. Finally the differences between the mean gains have

[1] Thorndike, E. L. *An introduction to the theory of mental and social measurements.* Second Edition. New York: Teachers College, 1913, p. 193.

been entered together with the reliabilities of these differences as computed by means of the formula for PE_{dis}.

Let us now return to our question "what, then, is the PE of the obtained difference 1.03?" The computations which we have briefly described yield 0.26 as the answer to this question. According to the accepted theory, this means that if a very large number of similar determinations—i.e., similar to the one which led to the difference 1.03—were made, half the determinations would lie between 1.03 plus 0.26 and 1.03 minus 0.26. In other words, it means that half the determinations would lie between 0.77 and 1.29. Furthermore, it is inferred that 82 percent of the determinations would fall between 1.03 increased by twice 0.26 and 1.03 diminished by twice 0.26, that is to say, 82 percent of these hypothetical results would fall between 0.51 and 1.55.

In this case it would be held that the chances are very great indeed that there is a real difference in favor of the large classes. If 82 percent of the hypothetical determinations fall between 0.51 and 1.55, the remaining 18 percent will be equally divided between determinations greater than and less than these amounts. It is therefore evident that according to the probable error theory, only 9 percent may be expected to be less than 0.51. In other words, we may say that in this case a real difference exists and that as to its size the chances are about ten to one that it will exceed one-half a T-unit.

The purpose of Table XXV is to exhibit for each grade and subject the difference between gains in large and small classes and the reliability of these differences as determined by the application of the formula $PE_{diff} = \sqrt{(PE_A)^2 + (PE_B)^2}$. The particular difference and probable error which we have been discussing—namely the difference of 1.03 and the probable error of 0.26—will be found in the last column of this table in connection with the second-grade entries.

It is customary in the discussion of reliability to represent an obtained measure as reliable if it is a certain number of times (say, four times) as great as its probable error. This way of determining whether or not a measure is reliable leaves much to

be desired. Reliability is relative, not absolute. Moreover, the degree of reliance to be placed on an obtained measure depends upon the use one wishes to make of it. If a measure is only twice as great rather than four times as great as its probable error, the chances are 4.5 to 1 that the true measure will lie between zero and twice the obtained measure. Under such circumstances, the chances are about 10 to 1 that the true measure is positive and greater than zero. This may be a sufficient degree of reliability for some purposes. On the other hand, a difference which is so small in relation to its probable error that it may be wiped out or reversed once out of ten trials, may not be sufficiently reliable for serious administrative action. Whether a difference which is so large as to be four times its probable error affords a basis of action will depend partly upon what the action is and partly upon the temperament of the person whose judgment determines the basis of action.

In the case of differences such as we are considering, the actual existence of the differences is not sufficient. For example, none save the most radical of educational enthusiasts would argue for small classes if the effectiveness in teaching them was only a very little more than the effectiveness in teaching larger classes. In other words, minute differences, even though their existence may be certain, will be generally thought to be insufficient to justify an expensive policy in order to secure them.[2]

The question, therefore, of the significance of the differences shown in Table XXV involves two points of view on neither of which is it possible to be dogmatic. The first is the size of the difference itself, taken of course, in reference to the size of gains from which it is derived. The second is the reliability of this difference, involving a consideration of both the size of the probable error and the size of the difference. We shall have to decide, for example, with reference to the difference of 1.03 for the second grade whether as a difference between 9.36,

[2] Of course, there may be other grounds for deciding upon a small-class organization. These reasons, whatever they may be, are unaffected by any of the arguments of this investigation. In other words, we are concerned merely with measurable differences in the product of teaching certain subjects.

the measure of gains in large classes, and 8.33, the measure of gains of the same pupils in small classes, it is large enough to be important. As has already been shown, it amounts to a 12 percent advantage for the large class. Again, we shall have to decide whether this difference (1.03) even if regarded as important is sufficiently likely to be found under varying circumstances and conditions in view of its probable error of 0.26. Our own judgment—which in this matter is probably no better than the reader's judgment—is that while the difference is relatively reliable in the sense that it probably exists it is too small to be of much importance as an argument for large classes in the second grade.

Table XXV shows differences in the other direction both for the fifth and seventh grades. The difference in the fifth grade for all subjects combined is 1.79 T-units, which is a 34 percent greater gain for the small classes than for the large. Here the significance of the advantage for the small class is considerably greater than the advantage for the large class found in the second grade. Moreover, the reliability of the difference in favor of the small class is relatively high, since the difference is six times its probable error (0.29).

In the seventh grade the difference based on all subjects combined is still greater in favor of the small classes. This difference (2.97) amounts to a 51 percent advantage for the small class. If the difference between sizes of classes had been greater in the seventh grade than we were able to have them, and if our classes in that grade had been more numerous, we should feel more confidence in the value of this figure. The difference in favor of the small classes in the seventh grade is, nevertheless, more than six times its probable error.

So far as the second, fifth, and seventh grades may be taken as typical of primary, middle, and upper grades in the elementary school, we have in the combination of the scores of pupils in these grades our best answers to the question of the relative advantage in measured teaching product of small and large classes. Table XXV shows (and practically the same figures are given in

Table XXIII) that according to the rotation method the mean gain per pupil for all elementary-school pupils in small classes was 7.93 T-units. The mean gain for these same pupils when taught by the same teachers in large classes was 6.86. This yields a difference in favor of the teaching in the small class of 1.07 T-units. The small class is thus favored over the large class by 16 percent.

There is a sense in which the difference of 1.07 is too small. It is obtained by finding the mean gains per pupil for all second-, fifth-, and seventh-grade children. This somewhat overweights the fifth grade where the number of pupils is large, and considerably underweights the seventh grade which has the fewest pupils but where the difference in favor of the small class is greatest. If we wished to give equal weight to each of the three grades, we should not obtain our mean gains from the original distributions but should merely average the three gains or more briefly the difference between gains for the three grades in question. The situation might be represented something like this: the typical second-grade child did better when he was taught in the large class by 1.03 T-units; the typical fifth-grade child did better when he was taught in the small class by 1.79 T-units; and the typical seventh-grade child likewise did better when taught in the small class by 2.97 T-units. If, therefore, we have these three children before us either physically or in our imagination and we wish to average their differences of gains, we shall, of course, find it if we divide the sum by three. This would yield 1.24 as the difference in favor of the small class instead of 1.07. This point of view gives equal weight to the second, fifth, and seventh grades.

As to the reliability of the difference of 1.07, the probable error of 0.18 indicates that the difference is practically six times its probable error. This would seem to be satisfactory. It means (so far as the probable error theory is itself justified) that if a large number of investigations like this one were made, half the determinations of differences derived in the same way would fall between 0.89 and 1.25. Between 0.71 and 1.43, 82 percent

of the determinations would lie. In fact, the chances would be about ten to one that a given determination would not fall below 0.71.

We have said nothing about the different subjects as they are represented in Table XXV. It will be noted that the large gains are represented in arithmetic and language, especially the latter. Both these gains are in favor of the small class and they account for the greater part of the general superiority shown in the final scores. In reading, which in point of time devoted to it, is incontestably the most important elementary-school subject, there is no clear superiority either for the large or small class. On the one hand, in comprehension the small class is slightly favored—so slightly, however, that one may disregard the figure, especially in view of the fact that its probable error is twice as great as itself. On the other hand, in reading rate the large class is definitely favored by a difference of 1.86 T-units or more than 22 percent with a probable error less than one-fourth the size of the difference. If, therefore, there is any definite conclusion with regard to reading it is that rate increases in the large classes while comprehension is unaffected.

Details are interesting here as elsewhere, and the reader is urged to study Table XXV. For example, comprehension of reading appears to be definitely favored by the larger classes in the second grade—a grade where the idea of comprehension greatly overshadows that of rate. The gain in favor of the large class is considerable and it is three times its probable error. In spelling, for which we have only second-grade records, the large class is favored although to an almost negligible extent. The difference is small in reference to the size of the gains (amounting to about four percent), and the probable error is actually greater than the difference. We cannot escape the question of the importance of these differences.

As to the final difference, whether a superiority of 16 percent is important in relation to the actual differences in class-size which our investigation deals with, is a question on which people will probably differ. To us it seems that the advantage

for the small class approaches the point where it should influence administrative action. This comes about, of course, through the favoring of the small classes in the fifth and seventh grades where, however, it must be admitted that—especially in the seventh grade—our material is not highly satisfactory. If, however, the extension of this investigation, which we hope to make, confirms these figures for the entire elementary school, we shall feel that financial pressure must be acute and that other considerations must be of unusual importance to permit the neglect of this difference as a basis of administrative action.

CHAPTER VII

A STUDY OF CLASS-SIZE IN HIGH SCHOOL

Method of Procedure

In this study the second-semester first-year pupils and the first-semester second-year pupils of four large Chicago high schools were used. These schools were Hyde Park, Lane Technical, Tilden Technical, and Harrison Technical. Complete data were secured from 134 high-school classes enrolling 2,821 pupils. The "Terman Group Test of Mental Ability" was given to these pupils,[1] who were then classified as follows:

1. When two or three classes in any subject could be taught by one teacher, the pupils were divided into large and small classes. For example, if there were 50 pupils taking first-year algebra, they were divided into two sections. The exact size of these sections was determined by the intelligence score of the pupils. The small group might contain from 15 to 20 pupils and the large group from 30 to 35 pupils. Some bright, medium, and dull pupils were put into each group so that the range of abilities was similar. Moreover, the pupils were placed in the sections so that the average intelligence score of one group was approximately the same as for the other group. One teacher then taught both of these sections.

2. When four sections in a subject could be taught by the same teacher, the pupils were first divided into two groups of about the same size, those obtaining low scores being placed in one group, and those obtaining high scores in the other. Each of these groups was then divided into two sections, one large and the other small, so that the average intelligence and variability of the sections were about the same. Thus a teacher taught two bright or two dull sections which were approximately equal in intelligence and variability, but different in size.

[1] A few were given the "Chicago Group Intelligence Test."

At the end of the semester the teacher gave the same examination to her classes. The term grades of the pupils in each class were also obtained. (Where letters were given for grades, they were converted into percents.) Since the teacher of each pair of large and small classes was the same, and since the pupils were of approximately equal intelligence and variability, it is possible to make a comparison between the different pairs of classes, and also between different pairs of individuals of equal intelligence, one individual having been instructed in a large class and the other in a small one.

Data from a Pair of English Classes

The average person would generally be satisfied to compare one pair of such classes and base his opinion upon the results. Or, it is quite possible that if he had two equally capable children in the high school under the same teacher, one child being in a small class and the other in a large class, he might be willing to base his estimate of the advantage of the different classes upon the relative success of the two children. Especially would this be the case if the result happened to correspond to his personal liking. To show the fallacy of making judgments on a small number of cases, data from a pair of English classes are presented in Table XXVI. These two classes were taught by the same teacher who had 40 pupils in the large class and 22 in the small class. (Only pupils with complete records are indicated in the table.) The first lines of the table give the ratings for two pupils, one of which (pupil 1) was taught in the large class, and the other (pupil A) in the small class. Pupil 1, taught in the large class, made 90 for the term mark and 91 on the examination; while pupil A, taught in the small class, secured a mark of only 80 on both the term's work and the examination. However, it will be seen that there was a difference of 45 points in the intelligence ratings for these two pupils in favor of the pupil taught in the large class. Pupil 6 of the large class, however, had the same intelligence rating as pupil A of the small class,

but made 8 points more on the term's work and 3 points more on the final examination. In this instance, where the pupils

TABLE XXVI. DATA FROM A PAIR OF ENGLISH CLASSES TAUGHT BY TEACHER NO. 18. (a) LARGE CLASS

Pupil	Rating of Pupil on		
	Intelligence	Term	Examination
1	175	90	91
2	144	80	76
3	142	90	90
4	137	80	75
5	135	80	80
6	130	88	83
7	123	85	80
8	124	80	80
9	123	78	70
10	120	90	88
11	117	80	75
12	113	89	90
13	109	80	78
14	107	78	85
15	106	80	88
16	105	84	83
17	101	85	80
18	97	90	88
19	96	75	70
20	96	90	90
21	93	80	70
22	91	80	84
23	91	85	80
24	90	75	70
25	89	75	65
26	88	78	67
27	88	80	78
28	88	80	78
29	88	68	52
30	87	75	70
31	83	85	85
32	80	75	70
33	77	80	78
34	77	75	70
35	69	85	80
36	63	68	59
37	45	65	78
38	44	75	70
Mean	101	80	77

TABLE XXVI. DATA FROM A PAIR OF ENGLISH CLASSES TAUGHT BY TEACHER NO. 18. (b) SMALL CLASS

Pupil	Rating of Pupil on		
	Intelligence	Term	Examination
A	130	80	80
B	124	75	76
C	123	90	87
D	120	80	78
E	101	75	70
F	93	85	85
G	83	78	80
H	80	80	85
I	77	78	75
J	77	80	85
K	73	80	75
L	69	80	84
M	54	80	84
N	51	80	80
Mean	90	80	80

were of nearly the same intelligence, the pupil in the large class made better marks than the pupil in the small class. However, if we consider pupil 21 and pupil F, we find that while each had an intelligence score of 93, the pupil in the small class made 5 more points on the term's work and 15 more points on the examination than the pupil of the large class.

The last line of the table presents the mean (average) rating for the two classes. The reader will note, (1) that the intelligence rating of the larger class was 11 points higher than that of the small class, (2) that the term mark in each case was 80 points, and (3) that the small class was superior to the large class on the examination mark by 3 points. The evidence shown for this pair of classes indicates that the large class, even if it has a much higher intelligence rating, does not secure better results than the small class. In other words, the data favor the small class. However, the next pair of classes examined might, as in the case of individual pupils, present evidence of a contrary tendency. The pair of classes taught by teacher No. 1 presented in Table XXVII is a good example. Here we have

a large class of 36 pupils and a small class of 12 pupils. The pupils of the large class obtained 2 points more on the intelligence rating than the pupils of the small class and surpassed them by 8 points on the term mark and 7 points on the examination mark.

From these comparisons it is plain that a small number of cases proves nothing. If, however, it is shown that for a large number of pairs of pupils or classes, the pupils in the smaller classes do considerably better work, then we shall have reasonable assurance that the small class is the most efficient unit.

TABLE XXVII. INTELLIGENCE SCORES AND RATINGS OF LARGE AND SMALL ENGLISH CLASSES

Teacher	Size		Average Intelligence Score		Average Term Mark		Average Examination Mark	
	Large Class	Small Class	Large Class	Small Class	Large Class	Small Class	Large Class	Small Class
3	34	17	54	51	77	79	[a]	[a]
11	35	21	56	56	71	73	71	76
1	36	12	59	57	86	78	85	78
9	41	20	77	73	76	77	76	75
10	38	20	90	90	76	77	80	78
5	44	18	92	88	73	72	74	74
2	37	17	95	100	69	71	[a]	[a]
8	32	19	95	88	72	76	69	74
16	39	21	96	96	76	83	77	83
15	36	21	97	98	75	76	[a]	[a]
6	38	19	98	100	74	81	72	80
18	40	22	101	90	80	80	77	80
21	38	26	106	99	78	82	80	84
4	40	18	112	114	82	81	78	85
13	41	21	116	110	75	73	75	72
14	38	21	120	122	88	87	88	87
19	32	24	127	144	77	76	82	81
20	37	24	130	127	80	79	81	83
22	38	26	130	135	71	77	52	69
17	41	22	131	120	77	78	78	75
12	43	21	136	142	80	83	78	81
7	35	19	139	137	77	82	80	83
All.....	38	20	103	102	77	78	77	79

[a] No data.

Justification of Teachers' Marks

At this point it is proper to raise the question as to whether we are justified in using teachers' marks as a measure of the

success of one pupil or class over another pupil or class. We admit that we should have preferred to use standardized tests had the available tests and school conditions warranted such a procedure. Since, however, pupils are promoted on the basis of teachers' marks, they do not seem to us to be inappropriate for the present study. In addition to this, the comparisons made are between the marks given by the same teacher to pupils in a large and a small class, and not between marks by different teachers. The literature on the unreliability of teachers' ratings has been concerned primarily with the rating of the same performance by different teachers. The comparability among marks given by the same teacher to pupils simultaneously taught and tested has not been seriously questioned.

For example, (Table XXVII), it is evident that teacher No. 1 and teacher No. 22 presided over entirely different situations. The latter taught pupils of a distinctly higher intelligence than the former and yet gave these more intelligent pupils very much lower examination and term marks. If we should compare the small class taught by teacher No. 1 with the large class taught by teacher No. 22, any effect due to class-size would be overshadowed by the difference in the teachers' standards of ratings. It is probable, however, that both teacher No. 1 and teacher No. 22 succeeded by their marks in discriminating between their own pupils; and for the purpose of the kind of comparison here made, this is all that is necessary.

Bright Pupils Do Not Work at Their Maximum Ability

An examination of Table XXVI reveals the fact that there is a decided waste of unused ability in both large and small classes for those pupils having high intelligence ratings. In the small class containing 14 pupils, the seven whose intelligence scores range from 83 to 130 make an average term mark of 80; while the seven having intelligence scores from 51 to 80 do exactly as well. Table XXVIII shows the standing of the seven brightest pupils and the seven dullest pupils according to the

intelligence rating in the large and small English classes taught by teacher No. 18. In the large class, seven pupils having the highest intelligence rating obtained an average intelligence score of 141, an average term mark of 85, and an average examination mark of 82; while the seven pupils having the lowest intelligence rating obtained an average intelligence score of 65, an average term mark of 75, and an average examination mark of 72. The

TABLE XXVIII. COMPARISON OF RATINGS OF BRIGHT AND DULL PUPILS IN LARGE AND SMALL CLASSES

	Large Class			Small Class		
	Intelligence Score	Term Rating	Examination Rating	Intelligence Score	Term Rating	Examination Score
Seven pupils having the highest intelligence rating	141	85	82	111	80	79
Seven pupils having the lowest intelligence rating	65	75	72	69	80	81
Difference between average ratings of bright and dull pupils	76	10	10	42	0	−2
Percent of difference between average ratings of bright and dull pupils	117	13	14	61	0	−.03

difference between the intelligence ratings of the bright and dull children is 76. Between their term marks the difference is 10; and between their examination marks it is also 10. In this large class the bright pupils made 117 percent larger intelligence scores, while their term and examination marks were only 13 or 14 percent higher.

In the small class the difference between the intelligence score of the seven bright pupils and the seven dull pupils was 42. The bright pupils made no higher term ratings than did the dull pupils, and they even made a poorer showing in their examination ratings. The surprising thing to note in connection with

this small class is that the seven brighter pupils made 61 percent better scores on the intelligence rating, but did no better work than the duller pupils as measured by their term and examination ratings.

These data suggest another problem in respect to class-size. Would classification on the basis of intelligence make any difference in the improvement made in large and small classes? The author attempted to organize large and small classes which contained only bright, average, or dull pupils. This was a difficult task since it required six or more sections of pupils in a given subject. It is reasonable to expect greater improvement from homogeneous groups of pupils than from heterogeneous groups. It is also possible that children can do more satisfactory work in larger classes if their intelligence ratings are about the same, than in smaller classes which are composed of very bright and very dull pupils. While these suppositions seem quite possible we have no experimental evidence to support them.

Comparison of Grades of All Pupils in Large and Small Classes

Results in English Classes

Table XXVII gives for each of 22 pairs of English classes the mean intelligence scores, mean term marks, and mean examination marks obtained from all the pupils in the different classes. The pairs of classes are arranged according to the intelligence scores of the large members of the pairs. The figures on the first line give data for one large and one small class taught by teacher No. 3. The large class contained 34 pupils and the small class, 17 pupils. The large class had an average intelligence rating of 54, against a rating of 51 for the small class. In spite of the slight advantage of the large class in the intelligence rating, the small class made an average term mark of 79, which was 2 points more than was made by the large class. The classes taught by teacher 11 had equal intelligence ratings and the small class surpassed the large class by 2 points on the term mark, and

by 5 points on the examination mark. On the other hand, in the classes taught by teachers 19 and 14, the small classes had the advantage in the intelligence rating but were surpassed by the paired large classes on both the term and examination marks. The last line of the table gives the average results for the 22 pairs of English classes. The average number of pupils in the large classes was almost twice as great as that in the small classes; yet the large classes were surpassed by the small classes by only 1 point on the term mark, and by 2 points on the examination mark.

TABLE XXIX. COMPARISON OF AVERAGE INTELLIGENT SCORES, AVERAGE TERM GRADES, AND AVERAGE EXAMINATION GRADES OF THE DIFFERENT SUBJECTS FOR 67 LARGE AND 67 SMALL CLASSES

Subject	Number of pairs of classes	Average Size		Average Intelligence Score		Average Term Grade		Average Examination Grade	
		Large	Small	Large	Small	Large	Small	Large	Small
English.....	22	38	20	103	102	77	78	77	79
Algebra.....	18	36	20	100	103	74	77	76	77
History.....	2	34	21	139	137	76	75	75	74
Geometry...	3	41	23	128	121	76	77	63	68
Stenography.	3	37	22	84	86	78	78	78	80
Latin.....	1	39	25	104	93	77	74	75	71
French.....	2	35	19	110	113	74	71	71	71
Typewriting.	3	39	25	83	88	73	77	80	83
Botany.....	4	28	18	126	122	74	74	72	73
Arithmetic...	3	39	23	90	90	75	75	63	64
General Science.....	6	36	22	110	110	75	77	75	76
Total.....	67	2457	1388	6987	6994	5055	5147	5019	5118
Average...	..	37	21	104	104	76	77	75	76

Results in Different High-School Subjects

A comparison of average intelligence scores and average examination and term marks for the different subjects is presented in Table XXIX. The first line of this table contains the average English marks derived from Table XXVII. In Table XXIX, as in other tables where classes in different subjects have been combined, the weighted mean (or weighted average) has been

used. For example, in Table XXIX there are 22 large English classes, 18 large algebra classes, two large history classes, etc. The weighted mean for the size of the large classes is obtained by multiplying the average sized English class by 22, the average sized algebra class by 18, the average sized history class by 2, and so forth, adding these results, and dividing by the total number of large classes. The total number of pupils in large classes was 2,457, and the total number of large classes was 67. Therefore, the weighted mean for the size of the large classes was 37. The average size of the small classes was 21. The 18 small algebra classes surpassed the 18 large algebra classes on the intelligence rating by 3 points, on the term mark by 3 points, and on the examination mark by 1 point.

The data for the 67 pairs of classes indicate that the average intelligence scores of the large and small classes were the same and that the small classes surpassed the large classes by 1 point on both the term and examination marks.

TABLE XXX. INTELLIGENCE AND RATINGS OF LARGE AND SMALL CLASSES WHERE THE LARGE CLASSES WERE AT LEAST TWICE THE SIZE OF THE SMALL CLASSES

Subject	Number of pairs of classes	Average Size		Average Intelligence Score		Average Term Grade		Average Examination Grade	
		Large	Small	Large	Small	Large	Small	Large	Small
English.....	8	39	16	90	91	77	78	77	79
Algebra.....	5	40	18	96	100	75	79	86	86
General Science.....	1	42	17	73	79	70	70	70	70
Botany.....	1	33	15	129	118	78	72	78	76
Stenography.	1	36	18	52	56	73	73	71	77
Total.....	16	623	284	1454	1481	1212	1234	1265	1285
Average...	..	39	18	91	93	76	77	79	80

Results Found When the Large Class Was Twice the Size of the Small Class

Table XXX shows the results when pairs of classes were selected in which the large class was at least twice as large as the

small class. The small classes exceeded the large classes by 2 points on the average intelligence score, by 1 point on the average term mark, and by 1 point on the average examination mark. Of these 32 classes we have 16 large classes with an average membership of 39 pupils, or a total of over 600 pupils doing about as well on the term and examination mark as the 16 small classes with an average enrollment of 18 pupils, or a total of less than 300 pupils. In other words, 16 teachers seem to be able according to their own ratings to teach 600 high-school pupils about as efficiently as they do one-half the number.

Summary

From the data thus far presented we may conclude that on the whole the small classes did slightly—very slightly—better on the term work and examination than the large classes with which they were paired.

The Amount of Superiority Which a Small Class Ought to Show

How much better should the small classes do in order to justify the added expense of instruction? In our opinion the superiority should be much greater than one or two points. On such a question, however, large differences of opinion might be expected. Doubtless some would insist that, if any increase could be shown in favor of small classes, the added cost would be justified; while others would probably require the small classes to surpass the large classes by 20 to 50 percentage points before the small classes would, in their opinion, be defensible. In order to obtain some opinions on this matter, the following question was asked at the meeting of the Central Ohio School Masters' Club held May 6, 1922:

If in the case of two high schools all factors were similar except class-size, and if in the first school the average class-size was 20, and in the second school it was 35, how much higher

average grades should the pupils obtain in the first school to justify the larger expense?

Table XXXI presents the 25 answers received. Nothing is so apparent as the lack of any sort of consistency. The range of judgments was from 0 to 50 points, the majority of them falling between 3 and 10. The average was 8.8 points. Since the majority of these men were superintendents and principals, they were, no doubt, influenced by the cost. The writer, at a later time, had occasion to present a paper before the mathematics section of the Thirty-fourth Educational Conference of the Academies and High Schools in relation with the University of Chicago. Before presenting the paper, the same question was asked but especially applied to high-school mathematics classes. The answers from the mathematics teachers ranged from 1 to 10 points, with the average at 7. All the teachers except one thought that the small classes should not be justified unless 5 or more points were in their favor.

The question which was raised at the Columbus and Chicago meetings was essentially the question raised by the data presented in this chapter. Despite the form of the question, it is clear that some of the people misunderstood it. The gentleman who gave his opinion as zero was interviewed. What he meant to indicate was that he should not expect any superiority for the school having classes averaging 20 pupils. In other words, he had lost sight of the question of the justification of small classes. He was of the opinion that within the limit of sizes mentioned in the question, no better work would be done in the small classes than in the large ones.

Again, the person who replied "50 percent" clearly did not understand the question. Under most circumstances there would not be room for the small classes to surpass the large classes by 50 points. Indeed this would only be true if the large classes were rated below 50. Probably what this man meant was that the small classes would surpass the large classes by 50 percent of the grade for the large classes. Even this, in general, is absurd. Teachers' marks seldom run below 70. A 50 percent

TABLE XXXI. JUDGMENT OF 25 SCHOOL MEN ON THE NUMBER OF POINTS BY WHICH SMALL CLASSES SHOULD SURPASS LARGE CLASSES IN ORDER TO BE JUSTIFIED

Number of Points.....	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	50	Average
Number of Judges.....	1	0	1	3	3	5	0	1	4	0	1	2	0	0	0	2	0	0	0	0	1	1	8.8

increase on a 70 percent rating will give us the impossible result of 105.

If we take the averages of 8.8 for the Ohio School Masters, and 7 for the mathematics teachers as indicating anything like a typical judgment of school people, we can say that the differences indicated in our tables are considerably smaller than will serve to justify the small classes. However, the lesson to be gained from these responses is that a large number of people—many of whom are, doubtless, taking a decided stand in favor of small classes—have never taken the trouble to reach a decision as to the amount of superiority which a small class ought to show.

CHAPTER VIII

THE EFFICIENCY OF LARGE AND SMALL CLASSES WHEN RECORDS OF PUPILS PAIRED ACCORDING TO EQUAL INTELLIGENCE SCORES ARE USED

Results of Paired Pupils

Because of administrative difficulties it was impossible to make up each pair of classes with exactly the same average of intelligence score. In each pair of classes, however, there were many pupils in the small class who had the same intelligence scores as pupils in the large class. Data for the pupils taught

TABLE XXXII. RATINGS OF PUPILS OF THE SAME INTELLIGENCE SCORE, CLASS OF TEACHER NO. 18 (ENGLISH)

Type of pupil	Large Class			Small Class		
	Pupil	Rating of Pupil		Pupil	Rating of Pupil	
		Intelligence	Term		Intelligence	Term
Bright	1	130	88	1′	130	80
	2	124	80	2′	124	75
	3	123	78	3′	123	90
	All	126	82	All	126	82
Average	4	120	90	4′	120	80
Dull	5	87	85	5′	87	75
	6	86	80	6′	86	85
	7	83	85	7′	83	78
	8	80	75	8′	80	80
	9	77	80	9′	77	78
	10	77	75	10′	77	80
	11	69	85	11′	69	80
	All	80	81	All	80	79

by teacher No. 18 are presented in Table XXXII. The first line shows the ratings of a pair of pupils having the same intelligence scores—namely, pupil No. 1 in the large class and pupil No. 1′

in the small class. In the case of this pair of pupils, the large class had an advantage of 8 points on the term rating. On the other hand, a comparison of the ratings made by pupils No. 6 and pupil No. 6′ shows the small class superior to the large class by 5 points. The pupils of these and of other classes were paired in the same manner.

Through this method of procedure it is possible to compare the term marks of pupils with identical intelligence ratings. Since the examination ratings enter into the term rating and since they differ but little from term ratings, they have been omitted in the remainder of the discussion. In addition to pairing the pupils according to equal intelligence scores we have divided the pupils roughly into three groups; the bright, the average, and the dull, on the basis of the intelligence test. The same method of grouping was used as was indicated in Chapter V; i.e., the pupils who were below or who just reached the median of the grade below the one to which they belonged were counted as dull, those who reached or surpassed the median of the next grade above the one to which they belonged were considered bright, and those who fell between these medians were regarded as average. For example, the median score for eighth-grade pupils is 87. Therefore any ninth-grade pupil who scored 87 or less was considered dull. Likewise, the tenth-grade median score is 122. Consequently, ninth-grade pupils making a score of 122 or more were considered bright. Ninth-grade pupils having scores ranging between 88 and 121, inclusive, were considered average. On this basis, pupils 1, 2, and 3 (Table XXXII) are bright, pupil 4 is average, and the rest dull. The bright pupils had identical term ratings in the large and small classes. The dull pupils made 81 points in the large class and 79 points in the small class. In this one class a small superiority is shown in favor of the large class.

Data From English Classes

Table XXXIII presents the data for 22 pairs of English classes in which a comparison is made between the bright, aver-

TABLE XXXIII. COMPARISON OF BRIGHT, AVERAGE, AND DULL ENGLISH PUPILS IN LARGE AND SMALL CLASSES

Teacher	Class-size		Number of Pairs of Pupils			Average Term Mark					
						Bright Pupils		Average Pupils		Dull Pupils	
	Large	Small	Bright	Average	Dull	Large	Small	Large	Small	Large	Small
3	34	17	.....	.....	15	.....	.....	.....	.....	77	79
11	35	21	.....	.....	16	.....	.....	.....	.....	73	74
1	36	12	.....	.....	12	.....	.....	.....	.....	87	78
9	41	20	.....	.....	9	.....	.....	.....	.....	70	76
10	38	20	.....	.....	19	.....	.....	.....	.....	76	77
5	44	18	.....	3	13	.....	.....	70	77	73	74
2	37	17	.....	7	10	.....	.....	73	73	68	70
8	32	19	.....	2	10	.....	.....	85	85	70	73
16	39	21	.....	.....	19	.....	.....	.....	.....	76	84
15	36	21	.....	5	11	.....	.....	72	80	75	75
6	38	19	.....	6	11	.....	.....	75	78	79	86
18	40	22	3	1	7	82	82	90	80	81	79
21	38	26	4	4	10	85	88	83	83	78	81
4	40	18	5	3	4	84	84	77	80	73	75
13	41	21	4	3	7	70	75	77	70	69	72
14	38	21	5	12	.....	86	86	86	88	.....	.....
19	32	24	8	4	3	84	74	75	80	67	70
20	37	24	8	10	.....	81	81	76	76	.....	.....
22	38	26	13	3	5	73	81	70	74	68	66
17	41	22	2	4	.....	75	80	73	75	.....	.....
12	43	21	14	4	.....	81	83	78	83	.....	.....
7	35	19	15	3	.....	75	81	74	84	.....	.....
All.....	38 (Mean)	20 (Mean)	81	74	181	80	82	77	80	75	77

age, and dull pupils in large and small classes. The classes taught by teacher No. 18 (details of which are given in Table XXXII) are shown on the twelfth line of the table. The pairs of classes are arranged in Table XXXIII in the order of their average intelligence ratings. The reader will notice that several of the classes (at least so far as the records are here considered) are composed entirely of dull pupils, for example, those taught by teacher No. 3. Also there are classes which contain only dull and average pupils (classes taught by teacher No. 5). Several of the bright classes, for example the classes taught by teacher No. 7, contain no dull pupils.

In these 22 pairs of English classes there were 81 pairs of bright pupils, 74 of average pupils, and 181 of dull pupils. The last line of the table indicates the ratings of the different types of pupils in the 22 English classes. The bright and the dull pupils made 2 points more in the small than in the large classes, while the average pupils made 3 points more. Thus, in the English classes the average pupils were slightly more benefited by the small classes than either the bright or dull pupils. Data similar to those given for the English classes are shown for all the subjects in Table XXXIV. The first line of the table gives the average ratings for all the pairs of English pupils as indicated in the last line of Table XXXIII. It also presents, in the last two columns, the average marks for the three groups of pupils combined. For all the English classes, we find that the pupils in the small classes made 2 points more on the term mark than the pupils in the large classes with whom they were paired.

Data From Classes in Different High-School Subjects

The second line of Table XXXIV presents the average ratings of the pupils in the algebra classes. Again we find that the average pupils are somewhat more benefited by the small classes, and also, on the average, that in general the pupils in the small classes do slightly better than those with whom they were paired. Likewise, in the case of the general science classes, the average

TABLE XXXIV. COMPARISON OF TERM GRADES OF BRIGHT, AVERAGE, AND DULL PUPILS IN LARGE AND SMALL CLASSES

Subject	Number of pairs of classes	Number of Pairs of Pupils			Average Class-size		Average Mark							
							Bright Pupils		Average Pupils		Dull Pupils		All Pupils	
		Bright	Average	Dull	Large class	Small class	Large	Small	Large	Small	Large	Small	Large	Small
English	22	81	74	181	38	20	80	82	77	80	75	77	77	79
Algebra	17	72	70	140	36	21	77	77	73	76	73	75	74	76
Latin	1	2	4	8	39	25	80	90	73	83	73	75	74	79
French	2	7	4	14	35	19	71	68	73	70	71	70	71	69
Typewriting	3	5	8	29	39	25	78	84	73	79	74	74	74	76
General Science	6	43	19	46	36	22	83	86	77	82	72	74	77	80
Stenography	3	7	11	26	37	22	80	80	78	84	75	80	77	81
Botany	4	12	15	9	28	18	73	78	74	71	70	68	73	73
Arithmetic	3	8	9	39	39	23	78	78	76	74	73	74	74	75
History	2	21	8	2	34	21	80	78	70	73	65	75	77	77
Geometry	3	26	10	13	41	23	77	80	79	74	71	73	76	77
All	66	284	232	507	37	21	79	80	75	78	74	76	76	78

pupils derived more benefit from instruction in the small classes than did the bright and dull pupils.

There were, in all, 66 pairs of classes.[1] In these classes there were 284 pairs of bright pupils, 232 pairs of average pupils, and 507 pairs of dull pupils, which makes a total of 1,023 pairs of pupils or 2,046 pupils. The average size of the large classes was 37, and the average size of the small classes was 21. The bright pupils in the small classes made 1 point more on the term mark than the pupils of the same intelligence in the large classes, the average pupils 3 points more, and the dull pupils 2 points more. Again, we find that all the types of pupils when taken together derived a little more benefit from the small than from the large classes, and that the average pupils obtained the greatest benefit.

Results of Pupils in Bright, Average, and Dull Classes

In the foregoing treatment we have compared bright, average, and dull pupils without reference to the general intellectual level of the classes in which they were taught. There are "bright" classes and there are "bright" pupils; but all the bright pupils are not in the bright classes. It is true, of course, that classes of the highest intellectual level will have a preponderance of bright pupils; but in the aggregate many pupils of equal ability will be found in "average" classes and a few even in "dull" classes. On the contrary, a few dull pupils will be found in classes whose intellectual level is on the whole entirely above their status. In other words, in respect to intelligence a pupil may be *in* a class but not *of* it.

Our data permit us to show the achievements of bright pupils whether they were taught in classes of high, medium, or low general status. The data also permit the same distinction to be made with respect to average pupils and dull pupils. A class,

[1] One pair of algebra classes was not considered in this part of the discussion because the pupils were given the "Chicago Group Intelligence Test," and consequently could not be compared directly on the point of intelligence with the pupils who had taken the "Terman Group Tests of Mental Ability."

TABLE XXXV. COMPARISON OF BRIGHT, AVERAGE, AND DULL ENGLISH PUPILS IN BRIGHT, AVERAGE, AND DULL CLASSES

Type of Classes	Number of Classes	Class-size		Number of Pairs of Pupils			Average Term Mark					
							Bright Pupils		Average Pupils		Dull Pupils	
		Large	Small	Bright	Average	Dull	Large	Small	Large	Small	Large	Small
Dull	12	38	19	3	24	152	82	82	75	78	76	77
Average	4	39	22	18	22	21	82	83	83	84	74	77
Bright	6	38	23	60	28	8	78	81	75	78	68	68

it is to be understood, is regarded as bright, average, or dull if the average intelligence score falls above, between, or below the critical scores which have already been described and used in the classification of individual pupils. Thus, of the English classes shown in Table XXXV, each of the first twelve (six large and six small classes) had average intelligence scores of less than 87; each of the next four classes scored between 88 and 121; while the last six classes scored 122 or higher.

The average ratings of the dull, average, and bright English classes are given in Table XXXV. There were 12 dull English classes. Of these, the large classes had an average membership of 38 and the small classes of 19. In these dull classes there were 3 pairs of bright pupils, 24 of average pupils, and 152 of dull pupils. The bright pupils made the same average term grade in the large and small classes. Still keeping our attention on the dull classes, we note that the 24 average pupils in the small classes made 3 points more than the corresponding pupils in the large classes, and that the dull pupils in the small classes made 1 point more than the pupils with whom they paired in the large classes.

In the case of the average classes—i.e., the classes having an average amount of intelligence—similar statements of difference for bright, average, and dull pupils reciting in these classes may be drawn from Table XXXV. The differences favor the small classes by 1 point for the bright and average, and 3 points in the case of the dull pupils. Finally, in the case of the "bright" classes, bright and average pupils did better in small than in large classes by 3 points. The dull pupils, however, in these bright classes did no better whether the classes were large or small. It is rather interesting to note that if a child is conspicuously out of place, the size of the class makes no difference. Dull pupils surrounded by bright pupils flounder to the same extent in large and small classes. Similarly, bright pupils in the presence of a preponderance of dull pupils stand out sharply and obtain the same ratings without regard to size of class—at least this is the suggestion of Table XXXV. Of course, the number of pupils

TABLE XXXVI. COMPARISON OF TERM GRADES OF BRIGHT, AVERAGE, AND DULL PUPILS IN BRIGHT, AVERAGE, AND DULL CLASSES

Type of Classes	Subject	Number of pairs of classes	Number of Pairs of Pupils			Average Class-size		Average Term Mark					
								Bright Pupils		Average Pupils		Dull Pupils	
			Bright	Average	Dull	Large	Small	Large	Small	Large	Small	Large	Small
Dull	English	12	3	24	152	38	19	82	82	75	78	76	77
	Algebra	10	5	37	129	37	20	91	89	76	79	73	75
	Latin	1	2	4	8	39	25	80	90	73	83	73	75
	Typewriting	2	.....	.....	29	40	27	.....	.....	.....	.....	74	74
	General Science	3	.....	3	46	36	21	.....	.....	73	83	72	74
	Stenography	2	.....	.....	26	37	22	.....	.....	.....	.....	75	80
	Arithmetic	2	.....	.....	39	38	24	.....	.....	.....	.....	73	74
	Geometry	1	1	5	9	45	23	95	95	86	86	75	72
	All	33	11	73	438	38	21	86	87	76	80	74	76
Average	English	4	18	22	21	39	22	82	83	83	84	74	77
	Algebra	3	17	6	11	31	19	77	75	68	71	71	74
	French	2	7	4	14	35	19	71	68	73	70	71	70
	Botany	3	10	10	7	28	17	74	79	74	71	71	68
	Arithmetic	1	8	9	.....	41	21	78	78	76	74	.....	.....
	History	1	7	2	1	30	17	79	80	70	80	70	60
	Geometry	1	13	4	2	40	24	72	76	73	62	63	67
	All	15	80	57	56	34	20	77	78	77	76	72	73
Bright	English	6	60	28	8	38	23	78	81	75	78	68	68
	Algebra	4	50	27	.....	39	25	76	76	70	74	.....	.....
	Typewriting	1	5	8	.....	39	21	78	84	73	79	.....	.....
	General Science	3	43	16	.....	36	22	83	86	78	82	.....	.....
	Stenography	1	7	11	.....	35	22	80	80	78	84	.....	.....
	Botany	1	2	5	2	28	20	70	70	74	70	65	65
	History	1	14	6	1	37	25	81	77	70	71	60	90
	Geometry	1	12	1	2	37	21	80	82	70	60	60	85
	All	18	193	102	13	37	23	79	81	74	78	66	72

who are out of place is necessarily small. There can only be a few dull pupils in a class rated as bright, and correspondingly, only a few bright pupils in a class rated as dull.

Table XXXVI presents similar data for the classes in all the subjects. The reader will observe that the figures for the English classes are the same as were given in Table XXXV. In the dull algebra classes instruction in small classes was most favorable to the average pupils, a little less so to the dull pupils, and least of all to the bright pupils. In the average algebra classes instruction in the small classes was equally favorable to the dull and average pupils, and much less so for the bright pupils. In the bright classes there were no dull pupils. In this group the small classes were more favorable to the average pupils than to the bright pupils.

A study of the different types of classes in all subjects combined and their relations to achievement of bright, average, and dull pupils reveals the following: (1) that in the dull classes, instruction in the small classes was most favorable to the average pupils, less so for the dull pupils, and least of all for the bright pupils, (2) that in the average classes, instruction in small classes did not favor to any appreciable degree either the bright, average or dull pupils, and (3) that in the bright classes, the small classes were of least benefit to the bright pupils, a little more beneficial to the average pupils, and most beneficial to the dull pupils. In the small classes, however, the number of pairs of pupils was only 13. This table also indicates that the small classes are most favorable to the average and dull pupils, and least favorable to the bright pupils.

As shown in Chapter VII, the amount of difference which should exist in favor of the small classes, in order that they might be justified, is a matter of opinion. The differences indicated in the analysis of the marks of bright, average, and dull pupils in bright, average, and dull classes in favor of the small classes, range from —1 to +6. From these data we have some foundation for the belief that small classes were more favorable to the average and dull pupils than to the bright pupils.

In comparing bright, average, and dull pupils in bright, average, and dull classes, the number of cases for each group is, of course, more limited than would be desirable. An investigation, therefore, of a much larger group would be necessary before definite conclusions could be drawn. In this connection it would also be advisable to have the classes so organized that each type of class, namely, the bright, average, and dull classes should be made up of distinctly one type of pupils with the exception of a few pupils of another type. For example, if the class was to be considered a dull class at least 90 percent of the pupils should be those who rate low on intelligence, and the other tenth of them should be either bright or average pupils. We should then have a definite situation in which a few bright pupils were grouped with dull pupils.

Conclusions and Suggestions for Further Study

The best information we have concerning the size of classes in the United States at present is that the median elementary class contains 38 pupils, the median junior high-school class, 28 pupils, and the senior high-school class, 25 pupils. This, of course, means that there are as many classes larger than these medians as there are classes smaller than the medians. The general opinion among educators is that the size of classes should be reduced; but, as between types of schools, the trend of opinion corresponds in general with the common practice of having the largest classes in the elementary school, smaller classes in the junior high school, and the smallest classes in the senior high school.

The problem of class-size has interested students to such an extent that several investigations of it have been attempted. These have produced results which, while conflicting, have tended to show that the size of a class is not a determining factor in efficiency of instruction.

These investigations did not, for the most part, use the modern principles and refined technique of educational experi-

mentation. An increasing degree of refinement, however, is noticeable among the more recent investigations. For example, it was only in the latter investigations that the attempt was made to "control" the experiment. In this study the author has hoped to show the possibility of using controlled experiments and other refined technique in the solution of the class-size problem.

In Chapters IV, V, and VI, the method relating to classes in elementary schools has been indicated. It was found on the whole that the small classes were a little more efficient than the large classes. And it was also found that for all elementary grades combined the small classes in general were of greatest advantage to the dull pupils, and of less advantage to the average and bright pupils. Likewise, in Chapter VII and VIII, the method used to measure class-size in the high school was set forth. The results for high-school classes show in general that the small classes have a very slight advantage over the large class. In practically all cases this advantage is less than two points on the usual percentage basis. Whether two points should be considered a large enough difference to justify the small classes is a question which must be decided from a subjective point of view. The opinions expressed by two groups of people in regard to the differences which should be found in order to justify classes of 20 as against classes of 35 averaged 5 and 7 points, respectively. We find also that in all high-school classes in the different subjects taken together the average pupils obtained greater benefit from the small classes than did the bright or dull pupils. A study of bright, average, and dull classes in all subjects combined indicated that the average pupils obtained more benefit in each type of class than did the bright or dull pupils.

We have already mentioned in Chapter III some of the limitations of the present study. For example, on account of administrative difficulties, it was impossible to make the differences in class-size in the elementary grades as large as would be desirable. Consequently it was impossible to make comparisons between classes of several different sizes. We are unable, therefore, to suggest the best size of class for these grades. Further

study should involve a comparison of a much larger number of classes so that many comparisons may be made; e.g., classes of less than 10 compared with those of 15–19, 20–24, 25–29, etc., and classes of 10–14 compared with those of 20–24, 25–29, etc. With such data in hand a correlation could be worked out between product and class-size. Moreover, further investigations involving class-size should be applied to all different types of public schools including elementary schools, junior high schools, senior high schools, and colleges.

In the case of elementary classes or other classes where the basis for comparison is the improvement made during a given period of time, the intelligence quotient might be a better unit for classifying the pupils than the intelligence score. Especially is this advisable where it is possible to obtain achievement tests which have duplicate forms. In high-school and university classes the special abilities of the pupils in a given subject should be used as partial criteria in classifying the pupils into equal-ability groups. For example, if we were attempting to pair two algebra classes, it might be well to know something about the pupil's ability in algebra if he has previously studied this subject, or at least to know something of his attainment in arithmetic. Another item which might well be taken into consideration in making equalized groups is effort. It is well-known that pupils of relatively high intelligence often do poor school work; and likewise, that pupils of relatively low mentality do a fair grade of school work. The factor of application or effort, no doubt, contributes to this condition. If instruments could be made available which would measure this factor, one additional variable could be taken care of.

One of the greatest difficulties in the class-size study is that of obtaining teachers of equal ability. The present study was limited to some extent due to the fact that the only method used for obtaining teachers of equal efficiency was to make comparisons directly between the results for one group of pupils and the results for another group taught by the same teacher. A preliminary investigation should be made which would provide a

large number of teachers of known ability to secure results. In this preliminary investigation, the classes, of course, should be of equal size. Then these teachers should teach classes of varying sizes, grouping classes together where teachers have equal ability and arranging these classes for each equal-ability group of teachers in the order of size. Having done this, it will be possible, on the basis of gains between an initial and final test, to correlate size with results. If there is a relationship (inverse or direct) we may answer the question "Is the line of relationship straight or partly so? In other words, are there regions where the figures for class-size have different relations to results?" We might also answer the question "What is the optimum size of classes (1) in general, (2) for teachers of different abilities, and (3) for children of different abilities?"

APPENDIX I

Bibliography

Almack, John C. "Class-size and Efficiency." The Journal of the National Educational Association, Vol. 12, No. 3, 1923.

American City Bureau, Report of Committee. "Know and Help your Schools. Second Report. An Interpretation of Inquiry No. 11. Relating to School Buildings and Grounds, Enrollment, and Size of Class in the National Survey of Urban Public Schools." New York: American City Bureau, March, 1921.

Bachman, Frank W. Final Report of the Committee on School Inquiry, 1911-13. Vol. 1, pp. 606-09.

Ballon, Frank W. "The Size of Sections (Classes) in the High Schools." High School Organization. Yonkers-on-Hudson: World Book Company, 1914, pp. 4-34.

Boston, School Committee. "Number of Pupils to a Teacher." Rules of the School Committee and Regulations of the Public Schools. Boston: Printing Department, 1910. pp. 72-77. (School Document No. 11, 1910.)

Boyer, Philip A. "Class Size and School Progress." Psychological Clinic, VIII, (May 1914) 82-90.

Bray, Samuel E. "Size of Class in Ordinary School." School Organization. London: W. B. Clive, 1911, pp. 62-74.

Breed, Frederick, S. and McCarthy, Grace D. "Size of Class and Efficiency of Teaching." School and Society, IV, (December 23, 1916) 965-71.

Brown, John F. "The Number of Pupils in Each Class." American High School. New York: The Macmillan Company, 1909, p. 161.

Chancellor, William E. "Classifying, Marking, Grading, and Promoting Pupils." Class Teaching and Management. New York: Harper and Brothers, 1910, pp. 199-232.

Cleveland, Educational Commission. "Number of Students in a High School Class, or Section for Recitation." Report 1906, pp. 16-17.

Cornman, Oliver P. "Size of Classes and School Progress." Psychological Clinic, III, (December 15, 1909) 206-12.

Davis, C. O. "The Size of Classes and the Teaching Load in the High Schools Accredited by the North Central Association." School Review, 31: 412-29, June, 1923.

Elliott, C. H. "Variations in the Achievements of Pupils." Teachers College, Columbia University, Contributions to Education, 1914.

"Facts on the Cost of Public Education and What They Mean." Bulletin 1, 1922, The National Education Association, Research Department, Washington, D. C.

Felter, William L. "The Size of Classes." Education Review, XLI, (April 1911) 418-20.

Harlan, C. L. "Relation of Size of Classes to Schoolroom Efficiency." Illinois State Teachers' Association. Proceedings, 1913, pp. 155-61.

Harlan, C. L. "Size of Class as a Factor in Schoolroom Efficiency." Educational Administration and Supervision, I. (March 1915) 195-209.

Merrill, Alice L. "Class-room Enrollment." Sierra Educational News, XIV, (June 1918) 342-43.

New York, Report of Committee, New York Teachers of Modern Language. "A Study of the Modern Language Situation and Recommendations Pertinent Thereto." Bulletin of High Points, New York City Schools, June 1921.

New York, Report of City Superintendent of Schools. July 13, 1913, pp. 41-44.

Perry, Arthur C. "The Size of the Class." Management of a City School. New York: The Macmillan Company, 1908, p. 181.

"Relation of Size of Class to School Efficiency," Urbana, Ill.: University of Illinois. V. 19, No. 45, Bulletin 10, July, 1922.

Rice, J. M. "Educational Research: A Test in Arithmetic." The Forum, XXXIV, 286-97.

Rice, J. M. "Educational Research: The Results of a Test in Language." The Forum, XXV, 269-93.

Rice, J. M. "Why Improved Educational Machinery Fails to Yield Better Products." The Forum, 36-96, July, 1904.

St. Louis, Board of Education. "Average Membership to Each Teacher." Sixty-first Annual Report for the Year ending June 30, 1915, pp. 360-63.

Smith, Eugene Randolph. "Size of Class as a Factor in Schoolroom Efficiency —A Reply." Educational Administration and Supervision, I. (June 1915) 384-89.

"Statistics of City School Systems 1917-18." Bulletin 1920, No. 24, U. S. Bureau of Education, Washington, D. C.

Strayer, George D. "Size of Class in Elementary Schools and High Schools." Some Problems in City School Administration. Yonkers-on-Hudson: World Book Company, 1916, pp. 94-97.

United States, Bureau of Education. "Size of Classes." A Study of the Colleges and High Schools in the North Central Association. Washington: Government Printing Office, 1915, pp. 47-53, Bulletin, 1915, No. 6.

VITA

Paul R. Stevenson was born November 26, 1886, at Cedar Rapids, Iowa. He received his early education in the public schools of Cedar Rapids and was graduated from Park College in 1909. From 1909 to 1911, he attended Art School in Cincinnati, Ohio. He did graduate work in education and psychology at the University of Cincinnati during 1912–14. During this time he also taught in a consolidated school and acted as assistant in the Department of Psychology. He received his Master's degree at the University of Cincinnati in 1914 and during the year 1914–15 taught mathematics in the high school at Cripple Creek, Colorado. He taught science in Highland College, Highland, Kansas, during the two-year period 1915–17. In 1917 he was appointed Dean of the University of Omaha and also Director of the Summer School. During the school year 1919–20, he did graduate work at Columbia University, receiving a Master's degree. In 1920 he accepted a position as Research Assistant in the Bureau of Educational Research, University of Illinois. Since 1921 he has been Research Associate in the Bureau of Educational Research, Ohio State University.